C-27
OCT 18 1938

AMERICAN JOURNAL OF MATHEMATICS

FOUNDED BY THE JOHNS HOPKINS UNIVERSITY

EDITED BY

E. T. BELL
CALIFORNIA INSTITUTE OF TECHNOLOGY

ABRAHAM COHEN
THE JOHNS HOPKINS UNIVERSITY

T. H. HILDEBRANDT
UNIVERSITY OF MICHIGAN

F. D. MURNAGHAN
THE JOHNS HOPKINS UNIVERSITY

J. F. RITT
COLUMBIA UNIVERSITY

WITH THE COÖPERATION OF

MARSTON MORSE	G. C. EVANS	OYSTEIN ORE
E. P. LANE	AUREL WINTNER	H. P. ROBERTSON
ALONZO CHURCH	GABRIEL SZEGÖ	M. H. STONE
L. R. FORD	R. L. WILDER	T. Y. THOMAS
OSCAR ZARISKI	R. D. JAMES	G. T. WHYBURN

PUBLISHED UNDER THE JOINT AUSPICES OF
THE JOHNS HOPKINS UNIVERSITY
AND
THE AMERICAN MATHEMATICAL SOCIETY

Memorial Number
George William Hill

Volume LX, Number 4

OCTOBER, 1938

THE JOHNS HOPKINS PRESS
BALTIMORE, MARYLAND
U. S. A.

THE AMERICAN JOURNAL OF MATHEMATICS will appear four times yearly.

The subscription price of the JOURNAL for the current volume is $7.50 (foreign postage 50 cents); single numbers $2.00.

A few complete sets of the JOURNAL remain on sale.

Papers intended for publication in the JOURNAL may be sent to any of the Editors.

Editorial communications may be sent to Dr. A. COHEN at The Johns Hopkins University.

Subscriptions to the JOURNAL and all business communications should be sent to THE JOHNS HOPKINS PRESS, BALTIMORE, MARYLAND, U. S. A.

Entered as second-class matter at the Baltimore, Maryland, Postoffice, acceptance for mailing at special rate of postage provided for in Section 1103, Act of October 3, 1917, Authorized on July 3, 1918.

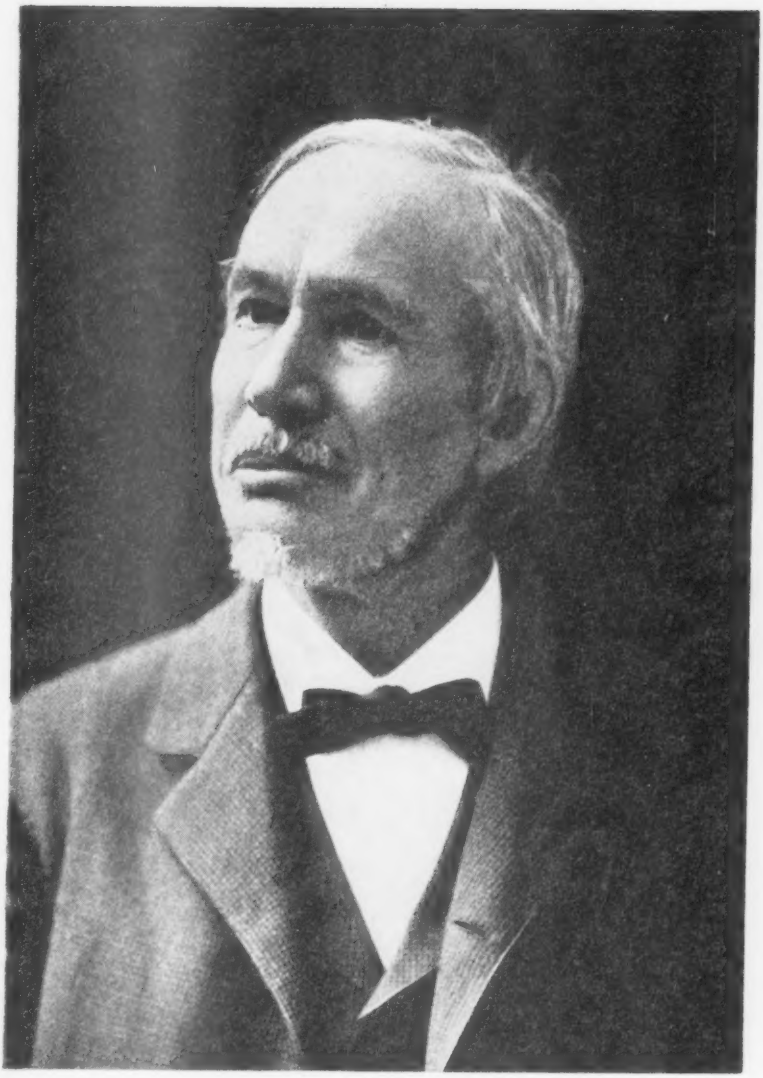
PRINTED IN THE UNITED STATES OF AMERICA
BY J. H. FURST COMPANY, BALTIMORE, MARYLAND

Memorial Number

George William Hill

GEORGE WILLIAM HILL

March 3, 1838 – April 16, 1914

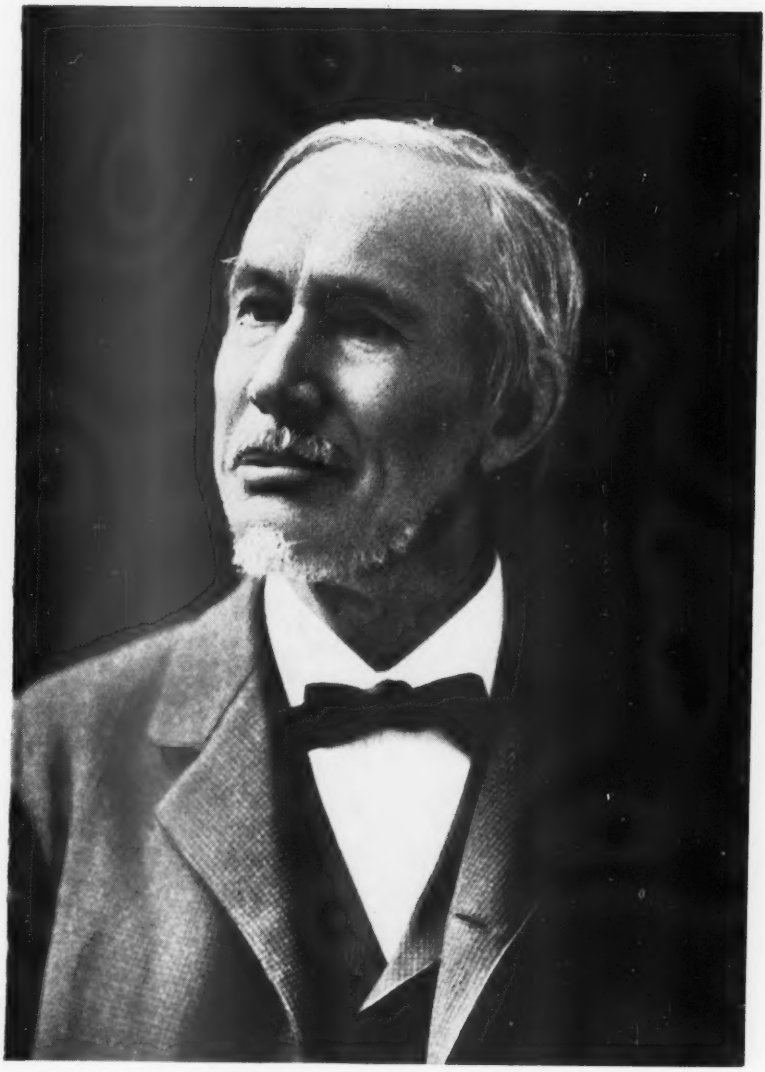


G. W. Hill.

UNIV. OF MICH.

GEORGE WILLIAM HILL

March 3, 1838 – April 16, 1914



G. W. Hill.

UNIV. OF MICH.

CONTENTS

PAGE

The equations of motion of the Moon. By E. W. Brown, . . . 785

Sur les avances du noeud et du périhélie d'une planète sous l'action d'une anneau circulaire. By Jean Chazy, ... 793

Die symmetrischen periodischen Bahnen des restringierten Dreikörperproblems in der Nachbarschaft eines kritischen Keplerkreises. By Ernst Hölder, ... 801

Symbolic dynamics. By Marston Morse and Gustav A. Hedlund, . 815

Periodic orbits in the restricted problem of three bodies in their relation to Hill's work on the motion of the Moon. By Elis Strömgren, 867

Zur Theorie der Dirichletschen Reihen. By Otto Toeplitz, . . 880

Mean motion. By Hermann Weyl, ... 889

The homogeneous chaos. By Norbert Wiener, ... 897

On Hill's periodic lunar orbit. By Aurel Wintner, ... 937

THE EQUATIONS OF MOTION OF THE MOON.

By Ernest W. Brown.*

1. In Volume 1 of this Journal, G. W. Hill published a paper entitled "Researches in the Lunar Theory" which has had far-reaching effects on the later development of the subject. From a study of it H. Poincaré conceived and developed the idea of the periodic orbit which in his hands became the basis for an extended study of the mathematics of the problem of three bodies. The calculation and classification of orbits of the type studied by Hill have formed the subject of many papers by various authors since his time. In my own case, it formed the starting point for a complete recalculation of the action of the sun on the motion of the moon. It is mainly the methods used in this last work that I am going to reconsider here.

It is not wise, I believe, to imagine that the work of a great master in his subject cannot be improved, especially when extended calculations are involved. As time goes on, new techniques are developed and in our own time calculating machines have come within the reach of everyone, so that a method previously rejected on account of the extended computation required may become easy and simple with but slight changes. My own experience in several cases leads me to think that it is worth while to return to the foundations even when they go so far back in time as those of Celestial Mechanics, and to examine whether valuable changes cannot be made.

2. The present paper is the result of such an examination. In my calculation, I adopted Hill's equations as they stood merely adding the terms in the disturbing function which complete accuracy demanded and adapting the solution of the equations to the additional terms. Two or more years ago it occurred to me to try the effect of referring the motion to axes which followed the mean position of the moon. I found later that Euler had had the same idea in his second theory. Hill had used axes which followed the mean position of the sun. There was a reason for this change as the coördinates would then take the form $a + x$, y where x, y are small and therefore expansions in powers of x, y were possible. What I did not expect, however, was that the equations for x, y could be put into a form such that the first gave x and then the second gave y with the same degree of approximation that had

* Professor Brown, who successfully undertook the great task of developing and applying Hill's theory of the Moon's motion, died on July 22, 1938, some days after revising the proof of the present paper. Editors.

been obtained before by the troublesome solution of two simultaneous equations. Further, the homogenous equation for x had fewer large terms present than the earlier equation which it superseded, that for y being substantially the same as before.

A further advantage in calculation is due to the fact that the lowest power of n' or m present is the second; in Hill's equations the first power occurs. Finally they have permitted me to find out the loss of accuracy due to small divisors which has occurred through the whole theory. This last feature is developed below and leads to the satisfactory result that the loss is considerably less than I had supposed. This fact has already appeared in the verification now under way.

When $e' = 0$, the time disappears explicitly from Hill's equations. This property has considerable advantage in theoretical researches and in the search for periodic orbits. But it is of no advantage in the calculation of satellite orbits. On the contrary it is a disadvantage in the latter case because the elimination of t deprives the equations of the advantages mentioned above for the new form.

3. The next step was an attempt to abbreviate the calculation of the various terms in the equations. In the so-called homogeneous form, if we neglect the small ratio of the distances, every term is a second degree term in x, y, z or their derivatives. As the coördinates consist of series of harmonic terms, by far the greatest part of the work consists of the multiplication of pairs of harmonic series. If then a technique could be developed for the multiplication of harmonic series by machinery practically the whole of the work of calculating the action of the sun on the moon could be done in perhaps a tenth of the time it originally required. I was fortunate in interesting my friend and former pupil, Professor W. J. Eckert of Columbia University, in this problem. He had already adapted several commercial machines of the Hollerith type to the solution of problems of celestial mechanics, and had developed a computing laboratory for their effective use. With the aid of the facilities of this laboratory we have been able in a year or two to test and extend calculations which took me nearly twenty years to carry out with the old-fashioned methods. These facts are mentioned because we are using the new equations in this verification.

4. The equations have also been used in another connection, namely, the exact calculation [1] of a portion of the motion of the perigee which had been neglected as insensible in the earlier work, but which I had found by another

[1] *Monthly Notices of the Royal Astronomical Society*, vol. 97 (1936), pp. 116-127.

method [2] must be large enough to account for most of the previously unexplained difference between the observed and theoretical values.

These equations appear to be the best to adopt for satellite motion when we can use the variation orbit as a first approximation and expand in powers of the eccentricities, inclination and ratio of the distances. For a satellite with a large eccentricity like Jupiter VIII which seems to require the ellipse as a first approximation, I hope to give elsewhere a method which appears to have considerable advantages over that which Professor Brouwer and I have used for this body and for which I was responsible.

The Non-Homogeneous Equations.

5. The equations of motion referred to an x-axis in the direction of the mean position of the moon, longitude $nt + \epsilon$, the y-axis perpendicular to it in a fixed plane and the z-axis perpendicular to this plane are:

$$5.1 \qquad \begin{cases} \ddot{x} - 2n\dot{y} + (1+x)\left(\dfrac{\mu}{a^3 r^3} - n^2\right) = \dfrac{\partial \Omega}{\partial x}, \\ \ddot{y} + 2n\dot{x} + \qquad y\left(\dfrac{\mu}{a^3 r^3} - n^2\right) = \dfrac{\partial \Omega}{\partial y}, \\ \ddot{z} \qquad + \qquad z \cdot \dfrac{\mu}{a^3 r^3} \qquad = \dfrac{\partial \Omega}{\partial z}, \end{cases}$$

where $a(1+x)$, ay, az are the coördinates of the moon, ar the distance and Ωa^2 the disturbing function due to the sun, supposed to move in an ellipse in the xy-plane.

When the moon is undisturbed and moves in a circular orbit in the plane of xy we have

$$x = y = z = 0, \qquad r = 1, \qquad a = a, \qquad \mu/a^3 = n^2.$$

Hence the factor in brackets in the first two equations is small.

If we add the integral of the second equation multiplied by $2n$ to the first we obtain an equation which may be written

$$\ddot{x} + (1+x)\left(\frac{\mu}{a^3 r^3} + 3n^2\right) = \frac{\partial \Omega}{\partial x} + 2n \int \left\{ \frac{\partial \Omega}{\partial y} - y\left(\frac{\mu}{a^3 r^3} - n^2\right) \right\} dt + \text{const.}$$

Since

$$1/r^3 = 1 - 3x + 6x^2 - \tfrac{3}{2}y^2 - \tfrac{3}{2}z^2 + \cdots$$

$\mu/a^3 r^3$ differs from n^2 by a small quantity of at least the first order and we know that Ω contains a factor of the second order; this equation therefore gives

$$\ddot{x} + n^2 x = \text{terms of the second order} + \text{const.}$$

---

[2] *Ibid.*, vol. 97 (1936), pp. 56-66.

provided the integral introduces no small divisors of the first order. The point of this equation is the fact that y or z are not present as first order terms. The equation can therefore be solved to give an approximation to x.

Since the last term on the left is of the second order, the second equation of 5. 1 evidently gives y when we know x. The third equation gives z in the same manner as the first gives x.

6. The actual procedure is so similar to that used in my Theory of the Motion of the Moon, that it is not necessary to repeat the various steps here. The chief difference is due to the fact that the various approximations to x, y are obtained separately instead of together by the solution of a pair of simultaneous equations as in the work mentioned. It may be noted that the $a(1+x)+aiy$ of this paper is the same as the $u\zeta^{-1}$ of the earlier work.

As the form of Ω differs somewhat I shall give it completely below, as far as it is needed.

We have [3]

$$6.1 \qquad \Omega a^2 = \frac{m' a^2}{a'^3}\left(\omega_2 + \frac{E-M}{E+M}\omega_3 + \frac{E^2-EM+M^2}{(E+M)^2}\omega_4 + \cdots\right),$$

where

$$\omega_2 = \frac{a'^3}{r'^3}\left(\tfrac{3}{2}r^2S^2 - r^2/2\right), \qquad \omega_3 = \frac{a'^4}{r'^4}\,\frac{a}{a'}\left(\tfrac{5}{2}r^3S^3 - \tfrac{3}{2}r^3S\right),$$

$$\omega_4 = \frac{a'^5}{r'^5}\,\frac{a^2}{a'^2}\left(3\tfrac{5}{8}S^4r^4 - 1\tfrac{5}{4}S^2r^4 + \tfrac{3}{8}r^4\right),$$

$$rr'S = (1+x)r'\cos(n't+\epsilon'-nt-\epsilon+E') + yr'\sin(n't+\epsilon'-nt-\epsilon+E'),$$

$$r^2 = (1+x)^2 + y^2 + z^2,$$

where $n't+\epsilon'$, E' are the mean longitude and equation of the center of the sun's orbit. In 6. 1, E, M are the masses of the earth and moon.

In order to expand these in powers of e' it will be found convenient to use complex quantities by putting

$$1+x+iy = u_1, \quad 1+x-iy = s_1, \quad nt+\epsilon-n't-\epsilon' = D, \quad e^{Di} = \zeta;$$

E', r' are functions of $l' = n't+\epsilon'-\omega'$, an angle which disappears when $e' = 0$. The expansions of the various functions of r', E' needed will then be found in T. M. M.[4]

We also put

$$m'+E+M = n'^2a'^3, \quad \frac{m'}{a'^3n^2} = \frac{n'^2}{n^2}(1-q) = m^2(1-q), \quad q = \frac{E+M}{E+M+m'}.$$

---

[3] Brown, *Lunar Theory*, Chap. I.

[4] These letters are used to denote my "Theory of the motion of the Moon," published in vols. 53 and 59 of the *Memoirs of the Royal Astronomical Society*.

If we are using numerical values, that of q can be inserted at the outset. Otherwise we put $q = 0$ and make a correction for it at the end.

The Homogeneous Equations.

7. Multiply the equations 5. 1 by $\dot{x}$, $\dot{y}$, $\dot{z}$ respectively, add and integrate. Noticing that t occurs in Ω, through the coördinates x, y, z as well as explicitly where it is shown and where it is present in E', r' and denoting this explicit occurrence by $\partial\Omega/\partial t$, we have

$$7.1 \quad \tfrac{1}{2}(\dot{x}^2+\dot{y}^2+\dot{z}^2) - \frac{\mu}{a^3 r} - \tfrac{1}{2}n^2\{(1+x)^2+y^2\} = \Omega - \int \frac{\partial\Omega}{\partial t}\,dt + \text{const.}$$

Now t occurs explicitly through $D = nt + \epsilon - n't - \epsilon'$ and through $n't + \epsilon' - \omega' = l'$. We have therefore

$$\frac{\partial\Omega}{\partial t} = \frac{\partial\Omega}{\partial S}\cdot\frac{\partial S}{\partial D}(n-n') + \frac{\partial\Omega}{\partial l'}n'$$

$$= \frac{\partial\Omega}{\partial S}\left\{-\frac{1+x}{r}\sin(D-E') - \frac{y}{r}\cos(D-E')\right\}(n-n') + \frac{\partial\Omega}{\partial l'}n'.$$

But

$$(1+x)\frac{\partial\Omega}{\partial y} - y\frac{\partial\Omega}{\partial x} = \frac{\partial\Omega}{\partial S}\left\{-\frac{1+x}{r}\sin(D-E') - \frac{y}{r}\cos(D-E')\right\}.$$

Hence

$$7.2 \quad \frac{\partial\Omega}{\partial t} = \left\{(1+x)\frac{\partial\Omega}{\partial y} - y\frac{\partial\Omega}{\partial x}\right\}(n-n') + \frac{\partial\Omega}{\partial l'}n' = \frac{dQ}{dt}(n-n') + \frac{\partial\Omega}{\partial l'}n'$$

this giving the definition of Q. The right-hand member of 7. 1 is therefore

$$\Omega - Q(n-n') - n'\int\frac{\partial\Omega}{\partial l'}\,dt + \text{const.}$$

The integral vanishes with e'.

Again, multiply equations 5. 1 by $1+x$, y, z respectively and add; we obtain a result which may be written

$$7.3 \quad \tfrac{1}{2}\frac{d^2}{dt^2}(r^2) - (\dot{x}^2+\dot{y}^2+\dot{z}^2) - 2n\{\dot{y}(1+x) - \dot{x}y\} + \frac{\mu}{a^3 r} - n^2\{(1+x)^2+y^2\} = r\frac{\partial\Omega}{\partial r},$$

where Ω in the right-hand member is supposed to be expressed in terms of r, S.

Adding this to 7. 1, we note that the negative power of r disappears and we get

$$7.4 \quad \tfrac{1}{2}\frac{d^2}{dt^2}(r^2) - \tfrac{1}{2}(\dot{x}^2+\dot{y}^2+\dot{z}^2) - 2n\{y(1+x) - \dot{x}y\} - \tfrac{3}{2}n^2\{(1+x)^2+y^2\}$$
$$= \Omega + r\frac{\partial\Omega}{\partial r} - Q(n-n') - n'\int\frac{\partial\Omega}{\partial l'}\,dt + \text{const.}$$

By multiplying the first two equations of 5. 1 by $-y$, $1+x$, and adding, we get a second equation independent of negative powers of r. This may be written after integration

$$7.5 \qquad \dot{y}(1+x) - \dot{x}y + n\{(1+x)^2 + y^2\} = Q + \text{const.}$$

Finally, multiplying this by $2n$ and adding the result to the previous equation we obtain

$$7.6 \quad \tfrac{1}{2}\frac{d^2}{dt^2}(r^2) + \tfrac{1}{2}n^2\{(1+x)^2+y^2\} - \tfrac{1}{2}(\dot{x}^2+\dot{y}^2+\dot{z}^2) = \Omega + r\frac{\partial\Omega}{\partial r} + Q(n+n') - n'\int\frac{\partial\Omega}{\partial l'}\,dt + \text{const.}$$

For calculation subtract from each side of the equation $-\frac{3}{4}m^2(x^2+y^2-2z^2)$. The right-hand member will then contain no term free from one or both of the explicit angles $2D$, l' except a small term depending on the square of the ratio of the parallaxes.

To get an equation for z, we subtract the first of 5. 1 multiplied by z from the third multiplied by $(1+x)$ and obtain

$$7.7 \quad (1+x)(\ddot{z}+n^2z) + z\left\{-\ddot{x}+2n\dot{y} - (1+x)\frac{\partial\Omega}{z\,dz} + \frac{\partial\Omega}{\partial x}\right\} = 0.$$

The equations 7. 6, 7. 5, 7. 7 are those needed. It will be noticed that if we express the coördinate $1+x$ by a single letter and confine Ω to the portion ω_2, the equations are homogeneous and of the second degree in x, y, z apart from the arbitrary constants.

8. Constants. The constants of the theory are found as in T. M. M., except perhaps the constant a. It is best to leave this to the end of the work and find it from the first of equations 5. 1, or 7. 3, the latter having the advantage of also giving $1/r$. In substituting the results obtained for x, y, z therein, I recall that here x is defined to contain no constant term. It consists entirely of cosines of multiples of the angles D, l, l', $2F$, while y consists of sines of the same angles. In the coördinate z the multiples of F present are odd, and the harmonics are sines.

9. Loss of accuracy due to small divisors. The approximate parameters are $m = 1/13$, $e = 1/20$, $\gamma = \sin\frac{1}{2}i = 1/22$, $e' = 1/60$, $a/a' = 1/400$. The effects of the first three of these generally will be discussed as though each of them was $1/10$.

Small divisors may arise in the x, z equations from terms which have a period of a month, and in the y equation from terms of periods long compared with a month. The latter will also arise in Q so that y may involve the square of the small divisor.

Small divisors are of two kinds. Both of them are independent of the mean motion of the moon. Those containing the mean motion of the sun will have the factor m and those independent of it the factor m^2. Examples of the former are $2D - l$, D, l', $2D - 2l$, and of the latter, $2F - 2l$, $D + l'$, $D + l' - l$. It will be noticed that each long period term is associated with two short period terms obtained by adding and subtracting l from its argument.

Suppose that we determine at any one step, as in T. M. M., the terms containing a given characteristic, i. e., those which have as a factor any given powers of e, γ, e', a/a'. In any such group there are either monthly or long period terms but not both. This is because the arguments in any such group always differ by even multiples of D, l, l', F. Thus as we proceed from lower to higher characteristics, short period and long period terms occur alternately, except as far as multiples of l' are concerned. But as $e' = 1/60$, and the main argument proceeds by decimal places, the additional factor $1/6$ takes care of accumulations and no special treatment is necessary.

Let us first take the cases of divisors which have m as a factor. Since Ω contains the explicit factor m^2 there is no loss of accuracy from these terms in Ω even when a double integration is involved. Thus the only terms which might cause a loss are those in $\dot{x}^2 + \dot{y}^2 + \dot{z}^2$ in the x-equation, those in $x\dot{y} - y\dot{x}$ in the y-equation, and the similar terms in the z-equation. But in such terms each of the coefficients is at the highest .1, so that no loss occurs when the product is divided by a number of the order .1.

Of the terms which have divisors with a factor m^2 the most important are the long period terms with arguments $2F - 2l$, $2D + 2l' - 2l$, $2D + 2l' - 2F$ and the associated monthly terms with arguments which differ from these by $\pm l$.

For the first, the term with the lowest order characteristic is the monthly term with the argument $2F - l$, and this is therefore the first to be computed. Now in the x-equation (7. 6), the terms on the right have the explicit factors m^2, so that no loss arises from these. In the second degree terms on the left as before, each factor is at most .1 so that one place is lost from these. But as the term arises mainly from the terms with argument l, $2F$ in x, y and as the term with argument $2F$ can be taken to an extra place without difficulty we can avoid this loss. There is also a combination from $\dot{z}^2$ with arguments $l - F$, $-F$, but the former can also be taken to an extra place. Hence we can avoid all loss for the term with argument $2F - l$.

The next in order is the long period term with argument $2F - 2l$. The additional part of the characteristic takes care of one decimal place, and the net loss is therefore one place.

In the succeeding monthly term there is no additional loss, because the

only place where this might occur is in $\dot{x}^2 + y^2$ and this involves the derivatives of the long period terms.

The second group of terms starts with the monthly term having the argument $2D + 2l' - l$. The only combination in the second order terms in which accuracy is lost is that with arguments $2D + 2l', l$. But the term with the former argument having a coefficient of the order .0001 can be taken to an additional place without trouble, so that loss of accuracy can be wholly avoided. In the succeeding long period term with argument $2D + 2l' - 2l$ we lose one place. With the monthly term $2D + 2l' - 3l$ no accuracy is lost.

The next term which needs consideration is that with argument $D + l' - l$ and having the characteristic $ee'a/a'$ so that it is of the fifth order. The preceding monthly term with argument $D + l'$ can be found without loss of accuracy because neither of the terms which might cause it is so great as .01. Hence proceeding as before we see that with this group only one place is lost.

Finally for the term $2D + 2l' - 2F$ which is the first of the group, the chief losses occur from the terms with argument $2D + 2l'$, $2F$ in x, y and $2D + 2l' - F, F$ in z. As $2D + 2l' - F$ in z is a monthly term with a divisor m^2 there would be a loss if we do not take the term $2D + 2l'$ in x to an extra place, which we can. Hence z can be obtained without loss and the term with argument $2D + 2l' - 2F$ with a loss of one place. There is no loss with the succeeding short period terms.

Incidentally it may be noticed that there is no loss of accuracy in the coördinate x due to a preceding short period or long period term, and that the only loss of accuracy is in y when the division is of order m^2.

Hence in the whole theory, it is not now difficult to see that the maximum loss can be reduced to two places of decimals. As I started in T. M. M. with 11 places of decimals, it is probable that as far as the small divisors are concerned the theory is correct to 0″001 in longitude. This result is considerably better than I had hoped. With the equations actually used it was not easy to limit the apparent losses in the manner shown above. They may, however, have been avoided in the course of the arithmetic without the computer being aware of the fact. The verification to which reference was made earlier has proceeded sufficiently far to enable me to state that amongst the terms containing even multiples of D, F — about half of the whole number — there appears to be no error in the earlier work greater than 0″01 in longitude, on the assumption that the transformation from rectangular to polar coördinates has been correctly computed. Two of the terms containing odd multiples of D apparently have errors of ″.03, ″.04. But it is not certain whether these are apparent only, some further work being necessary to decide the matter.

YALE UNIVERSITY OBSERVATORY.

SUR LES AVANCES DU NOEUD ET DU PÉRIHÉLIE D'UNE PLANÈTE SOUS L'ACTION D'UNE ANNEAU CIRCULAIRE.

Par JEAN CHAZY, Paris.

---

1. Dans mon Livre *La théorie de la Relativité et la Mécanique céleste*, pour discuter les différentes explications possibles de l'avance observée du périhélie de Mercure par rapport à la théorie des grosses planètes, j'ai étudié [1] l'action séculaire d'un anneau circulaire et homogène ayant pour centre le corps central, sur les mouvements du nœud et du périhélie d'une planète: et l'on sait, selon le principe de la méthode de Gauss, que l'action d'une autre planète animée d'un mouvement circulaire [2] équivaut à l'action d'un anneau circulaire et homogène. J'ai appliqué la méthode de calcul classique fondée sur les éléments osculateurs et le système différentiel de Lagrange.

Dans un Mémoire posthume [3] qui rassemble ses recherches de Mécanique céleste, Fatou a étudié plus généralement les mouvements presque circulaires d'un point matériel dans un champ de gravitation présentant un axe de révolution et un plan de symétrie perpendiculaires, et dérivant d'ailleurs d'un potentiel newtonien, voisin le plus souvent du potentiel du corps central. Fatou applique des méthodes directes dans lesquelles les parties principales des avances du nœud et du périhélie peuvent se déduire de deux équations différentielles linéaires du second ordre, dont les coefficients sont presque constants.

Je veux ici, dans le but de comparer les deux sortes de résultats, et développant d'ailleurs trois Notes antérieures,[4] reproduire ou du moins résumer le calcul de Fatou.

2. Nous déterminons le champ considéré par le potentiel

$$F(r, z) = \mu/r + R(r, z)$$

par unité de masse: μ désigne le coefficient attractif du corps central O, r la distance d'un point matériel P au point O, z la cote du point P comptée sur

---

[1] *Loc. cit.*, t. 1 (1928), pp. 195-203.

[2] Tous les mouvements circulaires et tous les anneaux circulaires que nous considérons dans cet Article ont pour centre le corps central.

[3] *Acta Astronomica*, Série a, vol. 2 (1931), pp. 101-163.

[4] *Comptes rendus des séances de l'Académie des Sciences*, Paris, t. 203 (1936), pp. 706, 844 et 981.

l'axe de révolution Oz à partir du point O ou du plan de symétrie II, et les deux fonctions F et R sont paires par rapport à la variable z. La fonction R est petite le plus souvent par rapport au terme μ/r. Les deux variables r et z conduisent à des formules plus simples que les coordonnées semi-polaires.

Considérées comme fonctions des coordonnées cartésiennes correspondantes, F et R sont des fonctions harmoniques, et les équations de Laplace relatives à ces fonctions se transforment en les deux équations

$$(1)\qquad \frac{\partial^2 F}{\partial r^2}+\frac{\partial^2 F}{\partial z^2}+\frac{2z}{r}\frac{\partial^2 F}{\partial r\partial z}+\frac{2}{r}\frac{\partial F}{\partial r}=0,\quad \frac{\partial^2 R}{\partial r^2}+\frac{\partial^2 R}{\partial z^2}+\frac{2z}{r}\frac{\partial^2 R}{\partial r\partial z}+\frac{2}{r}\frac{\partial R}{\partial r}=0,$$

équivalentes bien entendu l'une à l'autre; l'équation de Laplace relative à la différence

$$\frac{F-R}{\mu}=\frac{1}{r}$$

se réduit de même à l'identité à deux termes connue

$$\frac{\partial^2}{\partial r^2}\left(\frac{1}{r}\right)+\frac{2}{r}\frac{\partial}{\partial r}\left(\frac{1}{r}\right)=0.$$

Les deux dérivées $\partial^2 F/\partial z^2$, $\partial^2 R/\partial z^2$ ont évidemment la même valeur, et d'ailleurs Fatou a montré que leur valeur commune est négative dans les différentes applications.

Nous étudions des mouvements du point matériel P dont l'orbite est voisine d'une circonférence de centre O située dans le plan de symétrie II, et qui sont d'ailleurs de sens direct.

En coordonnées semi-polaires, soient r_1, θ et z $(r=\sqrt{r_1^2+z^2})$, Fatou considère dans le mouvement du point matériel P l'intégrale des aires

$$(2)\qquad r_1^2\theta'=C\qquad (C>0),$$

l'équation du mouvement relatif sur la projection du rayon vecteur sur le plan II

$$r_1''-r_1\theta'^2=\frac{\partial F}{\partial r}\frac{r_1}{r},$$

ou

$$(3)\qquad r_1''=\frac{C^2}{r_1^3}+\frac{\partial F}{\partial r}\frac{r_1}{r},$$

et enfin l'équation de projection sur Oz

$$(4)\qquad z''=\frac{\partial F}{\partial r}\frac{z}{r}+\frac{\partial F}{\partial z}.$$

En particulier soit un mouvement circulaire ayant lieu dans le plan de symétrie II, et où ρ et ω désignent le rayon et la vitesse angulaire : l'intégrale des aires (2) se réduit à

$$\rho^2\omega = C,$$

et montre que la vitesse angulaire ω est constante ; et l'équation (3) se réduit à l'équation

$$\frac{C^2}{\rho^3} + \frac{\partial F}{\partial \rho} = 0,$$

où $\partial F/\partial\rho$ désigne la valeur de la dérivée partielle $\partial F/\partial r$ pour $r = \rho$, $z = 0$. Et l'on tire de ces équations

$$\frac{1}{\rho}\frac{\partial F}{\partial \rho} = -\frac{C^2}{\rho^4} = -\omega^2. \tag{5}$$

3. En premier lieu, dans un mouvement qui tend vers le mouvement circulaire précédent, c'est-à-dire où la constante des aires C a exactement la même valeur que dans ce mouvement circulaire, et où le rayon vecteur r a des valeurs voisines du rayon ρ, calculons la durée de la révolution draconitique et l'avance du nœud. Dans le mouvement considéré, au second membre de l'équation (4), z est en facteur, puisque la dérivée $\partial F/\partial z$ s'annule avec z, d'une expression voisine de

$$\frac{1}{r}\frac{\partial F}{\partial r} + \frac{\partial^2 F}{\partial z^2},$$

et par suite de

$$\frac{1}{\rho}\frac{\partial F}{\partial \rho} - q,$$

ρ et ω étant définis par la constante C et par les équations (5), et $-q$ désignant la valeur commune des dérivées secondes $\partial^2 F/\partial z^2 = \partial^2 R/\partial z^2$ pour $r = \rho$, $z = 0$: q est positif puisque cette valeur commune est négative. Dès lors l'équation (4) peut s'écrire

$$z'' = -(\omega^2 + q)(1 + \epsilon)z, \tag{6}$$

ϵ désignant une quantité très petite.

Considérant z comme le seule fonction incomme, on sait que l'équation (6) s'intégre par deux quadratures ; Fatou étudie ces quadratures, mais, si nous nous bornons au terme principal de son calcul, nous pouvous dire que l'équation (6) est une équation linéaire en z dont les coefficients sont à peu près constants, et dont par suite les solutions sont voisines de fonctions périodiques de période $2\pi/\sqrt{\omega^2 + q}$.

En particulier la *durée de la révolution draconitique*, soit T_1, c'est-à-dire l'intervalle de temps compris entre deux instants consécutifs où la cote z s'annule en passant du négatif au positif, tend dans le mouvement considéré vers la même quantité

$$\lim T_1 = \frac{2\pi}{\sqrt{\omega^2 + q}}.$$

L'*avance du nœud*, soit $\delta\Omega$, c'est-à-dire la différence entre 2π et l'angle des rayons vecteurs correspondant aux deux instants qui précèdent, tend, puisque la vitesse angulaire tend elle-même vers ω, vers la valeur

$$\lim \delta\Omega = 2\pi\left(\frac{\omega}{\sqrt{\omega^2 + q}} - 1\right).$$

La fonction q étant positive en pratique, la valeur algébrique $\delta\Omega$ est négative: en fait, *dans le sens du mouvement, le nœud a, non une avance, mais un retard.*

4. En second lieu, concernant le terme principal de l'avance du périhélie, le calcul de Fatou peut se réduire à ceci. Soit un mouvement qui tend vers le mouvement circulaire considéré dans le plan Π, mais ayant lieu lui-même dans ce plan. L'équation du mouvement relatif sur le rayon vecteur, soit

$$r'' = \frac{C^2}{r^3} + \frac{\partial F}{\partial r},$$

peut, comme précédemment l'équation de projection sur Oz, être considérée comme une équation différentielle linéaire dont les coefficients sont à peu près constants. En effet le second membre s'annule pour $r = \rho$, d'après la définition de ρ et parce que la constante C a la même valeur que dans le mouvement circulaire limite, et d'autre part il a pour dérivée pour $r = \rho$

$$-\frac{3C^2}{\rho^4} + \frac{\partial^2 F}{\partial \rho^2} = -3\omega^2 + \frac{\partial^2 F}{\partial \rho^2},$$

quantité que nous supposons négative: donc l'équation peut s'écrire

$$r'' = \left(-3\omega^2 + \frac{\partial^2 F}{\partial \rho^2} + \epsilon_1\right)(r - \rho),$$

ϵ_1 désignant une quantité très petite. Donc les solutions de la dernière équation sont voisines de foncitons périodiques dont la period est la quantité

$$\frac{2\pi}{\sqrt{3\omega^2 - \frac{\partial^2 F}{\partial \rho^2}}}.$$

En particulier la *durée de la révolution anomalistique*, soit T_2, c'est-à-dire l'intervalle de temps compris entre deux instants consécutifs où le rayon vecteur r est minimum, tend vers la même quantité. On peut d'ailleurs transformer cette quantité, et écrire, compte tenu des équations (5) et de l'équation de Laplace mise sous la forme (1),

$$3\omega^2 - \frac{\partial^2 F}{\partial \rho^2} = \omega^2 - \frac{2}{\rho}\frac{\partial F}{\partial \rho} - \frac{\partial^2 F}{\partial \rho^2} = \omega^2 - q,$$

$-q$ désignant comme ci-dessus la valeur de la dérivée $\partial^2 F/\partial z^2$ pour $r = \rho$, $z = 0$. La révolution a donc pour durée limite

$$\lim T_2 = \frac{2\pi}{\sqrt{\omega^2 - q}};$$

et l'*avance du périhélie*, soit $\delta\varpi$, c'est-à-dire la différence entre 2π et l'angle de deux rayons vecteurs minima consécutifs, tend vers la valeur

$$\lim \delta\varpi = 2\pi\left(\frac{\omega}{\sqrt{\omega^2 - q}} - 1\right).$$

q étant possitif, la valeur algébrique $\delta\varpi$ est positive aussi : *tandis que le nœud a, dans le sens du mouvement, un retard, le périhélie au contraire a effectivement une avance.*

5. Nous allous montrer encore que dans les deux cas les résultats obtenus peuvent être déduits de la méthode classique de détermination des petits mouvements d'un système matériel dont la position dépend d'un nombre fini de paramètres ; et cette méthode donne même une extension du résultat obtenu dans le second cas.

En effet, dans le mouvement du point P, fixons encore la constante des aires C, éliminons l'angle polaire θ au moyen de l'intégrale des aires (2), et considérons dans le plan passant à chaque instant par l'axe Oz et par la position du point P, le mouvement relatif de ce point défini par le rayon vecteur r et la cote z, ou par les coordonnées cartésiennes r_1 et z.

La force relative dans ce mouvement est a priori la résultante de la force primitive, qui dérive de la fonction F, de la force centrifuge et de la force centrifuge composée correspondantes. Mais la force centrifuge est, au facteur -2 près, égale au produit vectoriel de la rotation instantanée du plan passant par Oz et par le point P, rotation dirigée suivant Oz, et de la vitesse relative du point P, vitesse située dans le même plan, qui est le plan du mouvement relatif que nous étudions. Donc la force centrifuge composée est perpendiculaire à ce plan, et par suite n'intervient pas dans le mouvement relatif

considéré, en particulier y a un travail nul dans un déplacement virtuel quelconque.

La force centrifuge n'intervient que par sa projection sur le plan du mouvement: cette projection se déduit de l'accélération normale du point coïncidant, c'est la force perpendiculaire à Oz, et de valeur algebrique $\theta'^2 r_1$ dans le sens où l'on s'éloigne de Oz; son travail dans un déplacement virtuel du point P, de composantes δr_1, δz, a pour expression

$$\theta'^2 r_1 \delta r_1 = \frac{C^2}{r_1^3}\delta r_1 = \delta\left(-\frac{C^2}{2r_1^2}\right) = \delta\left[-\frac{C^2}{2(r^2-z^2)}\right].$$

La force relative dérive ainsi, en fonction des variables r et z, de la fonction

$$U = -\frac{C^2}{2(r^2-z^2)} + F(r,z).$$

On peut dire aussi que l'équation des forces vives du mouvement primitif, soit

$$r'^2_1 + r_1^2\theta'^2 + z'^2 = 2F + 2h,$$

h désignant la constante des forces vives, se transforme dans le mouvement relatif où l'on a

$$r_1^2\theta'^2 = \frac{C^2}{r_1^2} = \frac{C^2}{r^2-z^2},$$

en l'équation

$$r'^2_1 + z'^2 = -\frac{C^2}{r^2-z^2} + 2F + 2h,$$

de sorte que le travail des deux forces centrifuges dans un mouvement relatif réel quelconque, est égal nécessairement, à une constante additive près, à la quantité $-[C^2/2(r^2-z^2)]$.

La valeur $z=0$, une valeur commune ρ du rayon vecteur r et de la distance r_1 à l'axe Oz, et la vitesse ω donnée par l'équation $\rho^2\omega = C$, déterminent un mouvement circulaire du point P, si ces valeurs correspondent dans le mouvement relatif considéré à une position d'équilibre relatif. Or la dérivée

$$\frac{\partial U}{\partial z} = -\frac{C^2 z}{(r^2-z^2)^2} + \frac{\partial F}{\partial z}$$

est nulle identiquement avec z, puisque F est fonction paire de z: donc les valeurs ρ, ω et $z=0$ doivent annuler l'expression de la dérivée

$$\frac{\partial U}{\partial r} = \frac{C^2 r}{(r^2 - z^2)^2} + \frac{\partial F}{\partial r},$$

soit

$$\frac{C^2}{\rho^3} + \frac{\partial F}{\partial \rho} = \omega^2 \rho + \frac{\partial F}{\partial \rho} = 0,$$

équations équivalentes aux équations (5).

Pour étudier les mouvements du point P voisins du mouvement circulaire correspondant à cet équilibre relatif, nous sommes ramenés à étudier parmi les mouvements relatifs considérés les petits mouvements voisins de la position d'équilibre obtenue. Appliquons la méthode classique avec $r - \rho$ et z comme variables indépendantes.

La force vive relative est

$$r'^2 + z'^2 = \frac{(rr' - zz')^2}{r^2 - z^2} + z'^2,$$

et se réduit à $r'^2 + z'^2$ si l'on fait $r = \rho$, $z = 0$ dans les coefficients de r'^2, $r'z'$, z'^2. De même, dans la fonction de force réduite, le terme rectangle est nul, car la dérivée $\partial^2 U/\partial r \partial z$ est nulle à cause de la parité de F et de U ; la dérivée $\partial^2 U/\partial z^2$ se réduit à la quantité négative

$$-\frac{C^2}{\rho^4} + \frac{\partial^2 F}{\partial z^2} = -\omega^2 - q,$$

selon la notation antérieure ; enfin la dérivée $\partial^2 U/\partial r^2$ se réduit à la quantité

$$-\frac{3C^2}{\rho^4} + \frac{\partial^2 F}{\partial \rho^2} = -3\omega^2 + \frac{\partial^2 F}{\partial \rho^2} = -\omega^2 + q,$$

que nous transformons comme dans le calcul antérieur, et il suffit que cette dernière quantité soit négative pour que la position d'équilibre considérée soit stable. La fonction de forces réduite est ainsi, après multiplication par 2,

$$-(\omega^2 - q)(r - \rho)^2 - (\omega^2 + q)z^2.$$

La méthode classique donne alors les deux équations

$$(7) \qquad \begin{cases} r'' = -(\omega^2 - q)(r - \rho), \\ z'' = -(\omega^2 + q)z. \end{cases}$$

Nous nous trouvons donc dans le cas simple où, au lieu de former un système différentiel linéaire d'ordre 4, les équations des petits mouvements sont deux équations différentielles linéaires du second ordre déterminant séparément chacune une variable. Les périodes fondamentales sont les deux quantités

$$\frac{2\pi}{\sqrt{\omega^2 - q}} \quad \text{et} \quad \frac{2\pi}{\sqrt{\omega^2 + q}}$$

considérées précédemment. Donc, selon la proposition générale, dans les mouvements du point P voisins de mouvements circulaires, le rayon vecteur r et la cote z sont représentés par des développements suivant les puissances des valeurs initiales $r_0 - \rho$, z_0, r'_0, z'_0. Les termes linéaires de ces développements sont donnés par les deux équations linéaires (7), ce qui met en évidence que les durées des révolutions anomalistique et draconitique tendent respectivement, quand les valeurs initiales tendent vers zéro, vers les deux quantités $2\pi/\sqrt{\omega^2 - q}$ et $2\pi/\sqrt{\omega^2 + q}$, d'où se déduisent immédiatement les limites des avances du périhélie et du nœud. [Remarquons que ces limites de la durée de la révolution anomalistique et de l'avance du nœud sont valables dans un mouvement quelconque voisin d'un mouvement circulaire, c'est-à-dire sans supposer comme précédemment l'orbite située tout entiere dans le plan de symétrie du champ de gravitation].

Les quatre valeurs limites étant obtenues ainsi de façons très simples, je les formerai dans un autre Article en introduisant les éléments osculateurs du mouvement et le système différentiel de Lagrange.

PARIS.

DIE SYMMETRISCHEN PERIODISCHEN BAHNEN DES RESTRINGIERTEN DREIKÖRPERPROBLEMS IN DER NACHBARSCHAFT EINES KRITISCHEN KEPLERKREISES.

Von Ernst Hölder.

Beim restringierten Dreikörperproblem ergibt sich im folgenden die Mannigfaltigkeit der (zur Syzygienachse symmetrischen) periodischen Bahnen (K) in der Nachbarschaft eines ("kritischen") Keplerkreises K^0 mit ganzzahligem Hillschen Periodenquotienten m als Sattelfläche

$$\text{(I)} \qquad \mu = \xi(\kappa - \tfrac{1}{2}a^{\frac{1}{2}}\xi^2 + \cdots)E(\kappa, \xi), \qquad E(0,0) = b \neq 0.$$

Diese wird von der Ebene verschwindenden Massenprozentsatzes μ im Punkte $\mu = \kappa = \xi = 0$ berührt und in den beiden Kurven geschnitten, die den Keplerkreisen $\xi = 0$ und den synodisch nach einem Umlauf geschlossenen Keplerellipsen entsprechen; κ ist der Zuwachs der Jacobischen Konstante von K gegenüber der von K^0, ξ im wesentlichen der $(m+1)$-te Fourierkoeffizient in der Kosinusentwicklung des Radiusvektor von K nach den Vielfachen der Elongation, ein Parameter, der sich für $\mu = 0$ näherungsweise auf die Exzentrizität reduziert; $a = (m/(m+1))^{2/3}$ ist der Radius von K^0.

Fixieren wir mit Wintner [1] den Massenprozentsatz $\mu > 0$, was der Strömgrenschen Gruppenfortsetzung entspricht, so ergeben sich auf der Lösungsfläche hyperbelartige Kurven, welche die von Poincaré [2] betonte analytische Fortsetzung der Lösungen einer natürlichen (isoenergetischen) Schar erster Sorte (über den kritischen Wert $\kappa = 0$ hinaus) in die Lösungen einer natür-

---

[1] A. Wintner, a) "Über eine Revision der Sortentheorie des restringierten Dreikörperproblems," *Ber. der Sächs. Akad. d. Wiss. zu Leipzig*, Bd. 72 (1930), S. 3-56; b) "Sortengenealogie, Hekubakomplex und Gruppenfortsetzung," *Mathematische Zeitschrift*, Bd. 34 (1931), S. 350-402. Auf diese Arbeiten stütze ich mich fortwährend, auch wegen ausführlicherer Literaturangaben sei auf sie verwiesen. Die von mir benutzte graphische Darstellung der Lösungsmannigfaltigkeit ist freilich eine andere als die auf der Massenschranke beruhende Wintnersche.

[2] H. Poincaré, "Les solutions périodiques et les planètes du type d'Hécube," *Bull. Astr.*, t. 19 (1902), S. 177-198, insbes. S. 185; die Figuren dort sind nur schematisch zu verstehen. Die zweite scheint mir auch nicht in allen Teilen die Tabelle von Brendel, "Theorie der Kleinen Planeten 1," *Abh. der Ges. d. Wiss. zu Göttingen*, Neue Folge 1, Nr. 2 (1898), S. 128, wiederzugeben.

lichen Schar zweiter Sorte vermitteln—Scharen, die nach den Untersuchungen von Levi-Civita [3] bezw. Birkhoff [4] feststehen.

Als Gegenstück werden auch die Verhältnisse in der Umgebung eines Keplerkreises K^0 mit rationalzahligem Periodenquotient $m = q/J$, $J > 0$, $(q, J) = 1$ wenigstens angedeutet, die sich ähnlich behandeln lassen. Dann gehört K^0 zwei verschiedenen Mannigfaltigkeiten von Lösungen an, welche die synodische Windungszahl 1 bezw. J haben, also zum ersten bezw. zweiten Poincaréschen Genre gehören und sich, wenn man wollte, in natürliche Scharen erster bezw. zweiter Sorte zusammenfassen ließen. Auch in jeder Ebene $\mu = \text{const.}$ existiert ein Doppelpunkt, durch den zwei verschiedene Lösungszweige hindurchgehen.

Die aus (I) approximativ folgende Beziehung

$$\xi = b^{-1}\mu\kappa^{-1},$$

im wesentlichen, hat sehr elegant Tisserand [5] hergeleitet — durch störungstheoretische Näherungsbetrachtungen, die sich aber wohl nicht so leicht zu einer strengen Methode für die Gewinnung des vollständigen analytischen Zusammenhangs ausgestalten ließen. Bei Tisserand ist zunächst nicht von periodischen Lösungen die Rede. Erst in einem kurzen Zusatz zur Darstellung seines Traité faßt er die ursprünglichen Entwicklungen von Radiusvektor und wahrer Länge als Fourrierreihen auf, deren Koeffizienten die in der störenden Masse linearen Approximationen der entsprechenden Koeffizienten einer periodischen Lösung sind. Diese hatte inzwischen, geleitet durch die Delaunaysche Mondtheorie, Hill [6] bemerkt und zugrundegelegt, als er im restringierten Dreikörperproblem Saturn—Titan—Hyperion, auf das übrigens auch Tisserand seine Theorie angewandt hatte, die Bewegung des Hyperion und die Titanmasse berechnete. Analytisch könnte die periodische Lösung mit derselben Methode der unendlich vielen Koeffizienten ermittelt werden, die Hill bei der Variationsbahn des Mondes entwickelt hat. Diese Methode hat neuerdings Wintner [7] mathematisch begründet und er hat auch die Anregung zu ihrer

---

[3] T. Levi-Civita, "Sopra alcuni criteri di instabilità," *Annali di Matematico* (3), vol. 5 (1905), pp. 221-308.

[4] G. D. Birkhoff, "The restricted problem of three bodies," *Rendiconti Palermo*, vol. 39 (1915), pp. 265-334.

[5] F. Tisserand, a) "Sur un cas remarquable du problème des perturbations," *Bull. Astr.*, t. 3 (1886), pp. 425-433; b) *Traité de la Mécanique Céleste*, t. 4 (1896), p. 113 ff., insbes. p. 116 f., sowie p. 491. Diese Untersuchung sowie die folgende von Hill hat Wintner, *loc. cit.* 1), a) hervorgehoben.

[6] G. W. Hill, "The motion of Hyperion and the mass of Titan" (1888), *Coll. Math. Works*, vol. 2 (1906), pp. 135-143.

[7] A. Wintner, a) "Zur Hillschen Theorie der Variation des Mondes," *Mathematische Zeitschrift*, Bd. 24 (1925), S. 259-265; b) "Über die Differentialgleichungen der

Anwendung auf das Problem der synodisch ganzzahligen Kommensurabilitäten gegeben, die von Martin [8] ausgeführt worden ist—allerdings nicht bis zu einem vollständigen Überblick über die Mannigfaltigkeit der Lösungen. Dazu muss man in der Rechnung noch einen Schritt weiter gehen und auch das Glied $\kappa\xi$ in (I) ermitteln. Ich tue das hier auf Grund der Lichtensteinschen [9] Theorie nichtlinearer Randwertprobleme, die mir [10] am nächsten liegt und deren ich vielleicht auch deshalb hier gedenken darf, weil sie ja eine grosse Analogie mit dem von Hill tatsächlich durchgeführten Verfahren hat: das Randwertproblem der symmetrischen Konjuktion und Opposition ist für symmetrische Bahnen mit dem der Periodizität gleichwertig, und die Berechnung der Störungen mittels mechanischer Quadraturen wird in der analytischen Theorie durch eine unter Zuhilfenahme der Greenschen Funktion gebildete Integrodifferentialgleichung ersetzt, die nach der Methode der sukzessiven Approximation gelöst wird.

1. Unter dem Einfluß des Newtonschen Gravitationsgesetzes sollen zwei Massenpunkte M_0 (Masse $1-\mu$, "Sonne") und M_1 (Masse μ, "Jupiter") umeinander Kreise beschreiben; der gemeinsame Radius der Relativbahnen sei gleich der Längeneinheit. Die Einheit der Zeit t sei so gewählt, daß die konstante Winkelgeschwindigkeit der Linie M_0M_1 gleich Eins ist.

In der Ebene der beiden Relativbahnen bewege sich ein Aufpunkt $P = P_\mu$ ("kleiner Planet"), welcher der Gravitationswirkung von M_0 und M_1 aus-

---

Himmelsmechanik," *Mathematische Annalen*, Bd. 96 (1926), S. 284-312; c) "Über die Existenz der Hillschen Mondbahn of maximum lunation und der Poincaréschen Schlingbahnen," *Mathematische Zeitschrift*, Bd. 28 (1928), S. 430-450; d) "Über die Konvergenzfragen der Mondtheorie," *Mathematische Zeitschrift*, Bd. 30 (1929), S. 211-227.

[8] M. Martin, "Upon the existence and non-existence of isoenergetic periodic perturbations of the undisturbed circular motions in the restricted problem of three bodies," *American Journal of Mathematics*, vol. 53 (1931), pp. 259-273.

[9] L. Lichtenstein, "Untersuchungen über die Gestalt der Himmelskörper, IV. Abh. Zur Maxwellschen Theorie der Saturnringe," *Mathematische Zeitschrift*, Bd. 17 (1923), S. 62-110; "V. Abh. Neue Beiträge zur Maxwellschen Theorie der Saturnringe," *Seeliger-Festschrift*, S. 200-225 (Berlin, 1925).

[10] Vgl. E. Hölder, a) "Mathematische Untersuchungen zur Himmelsmechanik," *Mathematische Zeitschrift*, Bd. 31 (1929), S. 197-256, insbes. S. 225-239. Eine Anwendung der Methode auf das Problem der synodisch ganzzahligen Kommensurabilitäten enthält meine Note b) "Die Verzweigungsgleichungen für die kritischen Kreise des restringierten Dreikörperproblems," *Ber. der Sächs. Akad. der Wiss.*, Bd. 83 (1931), S. 179-184. Aus dem dort gezeigten ausnahmslosen Nichtverschwinden des Koeffizienten von μ in der Verzweigungsgleichung folgerte ich damals mit Martin die Nichtexistenz einer von K^0 ausgehenden natürlichen Schar erster Sorte. Dieser Schluss muss aber ausdrücklich auf die von Hrn. Martin betrachteten Scharen eingeschränkt werden. Das vollständige Ergebnis wird erst hier gegeben und liefert eine positive Massenschranke auch für $\kappa = 0$, $\xi = 0$.

gesetzt ist, ohne diese Körper zu stören. Als Koordinaten von P nehmen wir die kartesischen Koordinaten x, y in dem rotierenden Achsenkreuz, dessen Ursprung in M_0 liegt und dessen x-Achse nach M_1 hin gerichtet ist.

Die Lagrangesche Funktion dieses restringierten Dreikörperproblems ist

$$(1) \quad \Lambda = \Lambda(x, y, \dot{x}, \dot{y}; \mu) = T + f + U,$$

wo

$$(2) \quad T = \tfrac{1}{2}(\dot{x}^2 + \dot{y}^2), \qquad f = x\dot{y} - y\dot{x},$$

$$(3) \quad U = \tfrac{1}{2}[(x-\mu)^2 + y^2] + (1-\mu)[x^2 + y^2]^{-\frac{1}{2}} + \mu[(x-1)^2 + y^2]^{-\frac{1}{2}}$$

und der Punkt die zeitliche Ableitung bezeichnet.

Da die Zeit t in Λ nicht explizit vorkommt, gilt das Jacobische Integral

$$(4) \quad T - U = k.$$

2. Man gewinnt nun die zur Jacobischen Konstanten k gehörigen Bahnkurven der Gestalt nach als Extremalen des geometrischen Variationsproblems $\delta \int \Lambda^* d\tau = 0$ mit der Maupertuisschen Wirkungsfunktion

$$(5) \quad \Lambda^* = [2(U + k) \cdot 2T^*]^{\frac{1}{2}} + f^*,$$

wobei T^*, f^* aus T, f entstehen, wenn man die Geschwindigkeitskomponenten $\dot{x}$, $\dot{y}$ durch die Ableitungen x', y' nach dem Parameter τ ersetzt. Dieser ist aus Homogenitätsgründen willkürlich und soll als der durch

$$(6) \quad x = r\cos\tau, \qquad y = r\sin\tau$$

definierte Polarwinkel τ ("Elongation") genommen werden, der also jetzt unabhängige Variable ist, während die gesuchte Funktion der Radiusvektor $r = r(\tau)$ ist. Es wird dann

$$(7) \quad \Lambda^* = [2(U + k)(r'^2 + r^2)]^{\frac{1}{2}} + r^2.$$

Der zeitliche Ablauf $\tau = \tau(t)$ der Bewegung, der uns hier nicht weiter interessiert, bestimmt sich wegen (4) hinterher durch eine Quadratur aus

$$(8) \quad \left(\frac{d\tau}{dt}\right)^2 = \frac{2(U + k)}{r'^2 + r^2}.$$

3. Wir untersuchen hier die Bahnen K, für die der Radiusvektor eine in τ mod 2π periodische Funktion $r(\tau)$ ist und die in der Nachbarschaft der zum Zweikörperproblem $\mu = 0$ gehörigen Keplerschen Kreisbahn

$$(9) \quad K^0: \quad x = a\cos\tau, \quad y = a\sin\tau$$

von P_0 liegen. Die Bahn K^0 hat die Jacobische Konstante, vgl. (51),

$$k^0 = -(2a)^{-1} - a^{\frac{1}{2}} \tag{10}$$

und die mittlere Bewegung n mit $n^2a^3 = 1$.

Der Charakter der gesuchten Lösungsmannigfaltigkeit (K) hängt wesentlich davon ab, ob für diese Ausgangslösung K^0 der Hillsche Periodenquotient

$$m = 1/(n-1) \tag{11}$$

eine ganze Zahl ist (singulärer, kritischer Fall) oder nicht (regulärer Fall). Die mittlere Bewegung ist jetzt $n = 1 + 1/m$.

Übrigens sind aus dynamischen Gründen die Werte

$$m = -1,\ 0,\ \infty,\ -\tfrac{1}{2} \tag{12}$$

auszuschließen; sie entsprechen bezw. den Grenzfällen, daß K^0 ins Unendliche rückt, in die Sonne hineinfällt, mit der siderisch direkten oder retrograden Jupiterbahn identisch ist.

Eine eingehendere Betrachtung bedarf nur der singuläre Fall; er kann nur für einen siderisch direkten Planeten eintreten. Denn die oberen und unteren retrograden Kreise gehören zu

$$-1 < m < -\tfrac{1}{2} \text{ bezw. } -\tfrac{1}{2} < m < 0, \tag{13}$$

können also niemals kritisch werden.

4. Für die zu untersuchenden Bahnen (K) sei der Radiusvektor

$$r = a(1+\zeta) \tag{14}$$

gesetzt; $\zeta = \zeta(\tau)$ muß die aus der Grundfunktion Λ^*, Formel (7), abgeleitete Lagrangesche Gleichung für ζ befriedigen. Daß K in einer Nachbarschaft von K^0 liegt, bedeutet, daß für K der Massenprozentsatz μ und die Differenz $\kappa = k - k^0$ der Jacobischen Konstanten absolut genommen kleine Werte besitzen und daß auch $\zeta(\tau)$ und seine Ableitung $\zeta'(\tau)$ dem Betrag nach unterhalb gewisser Schranken bleiben. Wir entwickeln am besten gleich Λ^*, wo auch $U = U(\tau, \zeta; \mu, \kappa)$ aufzufassen ist, nach Potenzen von μ, κ, ζ, ζ'. Dies wird für jeden einzelnen Wert von a gemacht.

Zum Zwecke der Entwicklung schreiben wir ausführlich

$$\begin{aligned} 2(U+k) = {} & a^2(1+\zeta)^2 - 2\mu a(1+\zeta)\cos\tau + \mu^2 - (2a^{\frac{1}{2}} + a^{-1}) + 2\kappa \\ & + 2(1-\mu)a^{-1}(1+\zeta)^{-1} \\ & + 2\mu[a^2(1+\zeta)^2 - 2a(1+\zeta)\cos\tau + 1]^{-\frac{1}{2}} \end{aligned} \tag{15}$$

und bekommen für die Lagrangesche Funktion den Ausdruck

(16) $\Lambda^* = a[(V + 2W\mu + 2\kappa + \mu^2)(1 + 2\zeta + \zeta^2 + \zeta'^2)]^{\frac{1}{2}} + a^2(1+\zeta)^2$

mit

(17) $V = a^2(1+\zeta)^2 + 2a^{-1}(1+\zeta)^{-1} - (2a^{\frac{1}{2}} + a^{-1}),$

(18) $W = [d^2 + 2a(a - \cos\tau)\zeta + a^2\zeta^2]^{-\frac{1}{2}} - a(1+\zeta)\cos\tau - a^{-1}(1+\zeta)^{-1};$

hier ist

(19) $$d^2 = 1 - 2a\cos\tau + a^2$$

und, was sofort benutzt wird, (11) zufolge

(20) $$a^{-\frac{1}{2}} - a = am^{-1}, \qquad a^{-3} = (m+1)^2 m^{-2}.$$

In der Entwicklung von V und W brauchen wir, wie sich herausstellen wird, die Anfangsterme bis zum dritten bezw. ersten Grad in ζ:

(21) $$\begin{aligned} V_3 &= (a - a^{-\frac{1}{2}})^2 + 2(a^2 - a^{-1})\zeta + (a^2 + 2a^{-1})\zeta^2 - 2a^{-1}\zeta^3 \\ &= a^2m^{-2}\{1 + 2(1 - a^{-3})m^2\zeta + (1 + 2a^{-3})m^2\zeta^2 - 2a^{-3}m^2\zeta^3\} \\ &= a^2m^{-2}\{1 - 2(2m+1)\zeta + (3m^2 + 4m + 2)\zeta^2 - 2(m+1)^2\zeta^3\} \end{aligned}$$

und

(22) $$W_1 = d^{-1} - a^{-1} - a\cos\tau + (a^{-1} + (\cos\tau - a)ad^{-3} - a\cos\tau)\zeta.$$

Wir erhalten daher

(23) $$\begin{aligned} \Lambda^* &= a(V + 2W\mu + 2\kappa + \mu^2)^{\frac{1}{2}}(1 + \zeta + \tfrac{1}{2}\zeta'^2 - \tfrac{1}{2}\zeta\zeta'^2) + a^2(1+\zeta)^2 + \Lambda^{(1)} \\ &= aV_3^{\frac{1}{2}}(1 + \zeta + \tfrac{1}{2}\zeta'^2 - \tfrac{1}{2}\zeta\zeta'^2) + a^2(1+\zeta)^2 \\ &\quad + aV_1^{-\frac{1}{2}}W_1(1+\zeta)\mu + aV_2^{-\frac{1}{2}}(1 + \zeta + \tfrac{1}{2}\zeta'^2)\kappa - \tfrac{1}{2}aV_1^{-3/2}(1+\zeta)\kappa^2 + \Lambda^{(2)}. \end{aligned}$$

Dabei ist

(24) $$\begin{aligned} V_3^{\frac{1}{2}} = am^{-1}\{1 &- (2m+1)\zeta - \tfrac{1}{2}(m^2-1)\zeta^2 \\ &- \tfrac{1}{2}(m+1)(2m^2 + m + 1)\zeta^3 + \cdots\}, \end{aligned}$$

(25) $$V_2^{-\frac{1}{2}} = a^{-1}m\{1 + (2m+1)\zeta + \tfrac{1}{2}(9m^2 + 8m + 1)\zeta^2 + \cdots\}.$$

$V_1^{-\frac{1}{2}}$ ist der Abschnitt davon bis zu dem Glied mit ζ und

(26) $$\begin{aligned} aV_1^{-\frac{1}{2}}W_1 = m\{d^{-1} - a^{-1} - a\cos\tau &+ [a^{-1} + ad^{-3}(\cos\tau - a) - a\cos\tau \\ &+ (2m+1)(d^{-1} - a^{-1} - a\cos\tau)]\zeta + \cdots\}; \end{aligned}$$

$\Lambda^{(1)}$, $\Lambda^{(2)}$ und nachher $\Lambda^{(3)}$ enthalten außer einigen für die Lagrangesche Gleichung belanglosen, von ζ, ζ' unabhängigen Gliedern solche Glieder dritter

Ordnung in κ, μ, ζ, ζ', die den Faktor μ enthalten, ferner Glieder vierter und höherer Ordnung. Schließlich wird

$$
\begin{aligned}
(27)\quad a^{-2}m\Lambda^* &= \tfrac{1}{2}\zeta'^2 - \tfrac{1}{2}(m+1)^2\zeta^2 - (m+1)\zeta\zeta'^2 - m(m+1)^2\zeta^3 \\
&\quad + a^{-2}m^2\{2(m+1)\zeta\kappa - a^{-2}m^2(3m+2)\zeta\kappa^2 \\
&\quad + \tfrac{3}{2}(m+1)(3m+1)\zeta^2\kappa + \tfrac{1}{2}\zeta'^2\kappa \\
&\quad + [2(m+1)d^{-1} - (2m+3)a\cos\tau \\
&\quad + ad^{-3}\cos\tau - a^2d^{-3} - (2m+1)a^{-1}]\zeta\mu\} + \Lambda^{(3)} \\
&= \tfrac{1}{2}\zeta'^2 - \tfrac{1}{2}(m+1)^2\zeta^2 + \Sigma B_{ijkl}\kappa^i\mu^j\zeta^k\zeta'^l
\end{aligned}
$$

abgekürzt.

5. Die Lagrangesche Gleichung ist

$$
\begin{aligned}
(28)\quad L(\zeta) \equiv \zeta'' + (m+1)^2\zeta &= \Sigma\{kB_{ijkl}\kappa^i\mu^j\zeta^{k-1}\zeta'^l - l\kappa^i\mu^j(B_{ijkl}\zeta^k\zeta'^{l-1})'\} \\
&= -\Sigma\{k(l-1)B_{ijkl}\kappa^i\mu^j\zeta^{k-1}\zeta'^l + l(l-1)B_{ijkl}\kappa^i\mu^j\zeta^k\zeta'^{l-2}\zeta'' \\
&\quad + lB'_{ijkl}\kappa^i\mu^j\zeta^k\zeta'^{l-1}\} \\
&= -B_{0012}\zeta'^2 + (3B_{0030}\zeta^2 - 2B_{0012}\zeta\zeta'') + B_{1010}\kappa + B_{2010}\kappa^2 \\
&\quad + (2B_{1020}\kappa\zeta - 2B_{1002}\kappa\zeta'') + B_{0110}\mu + \Psi^{(1)} = \Psi.
\end{aligned}
$$

Hier enthält $\Psi^{(1)}$ die Glieder zweiter Ordnung, soweit sie μ als Faktor haben, sowie die Glieder dritter und höherer Ordnung.

Die Ermittlung der in τ mod 2π periodischen Lösungen der Lagrangeschen Differentialgleichung (28) führt im regulären Falle $m \neq 1, \pm 2, \cdots$ mittels der Greenschen Funktion von $L(\zeta)$ zu einer nichtlinearen Integrodifferentialgleichung, deren Lösung $\zeta = \zeta(\tau; \kappa, \mu)$ sich für kleine Beträge von κ, μ durch sukzessive Approximationen ergibt und auch die Lagrangesche Gleichung (28) erfüllt und periodisch ist.

Im kritischen Fall $m = 1, \pm 2, \pm 3, \cdots$ hat $L(Z) = 0$ aber die periodischen Nullösungen $Z = \cos p\tau$, $\sin p\tau$ mit $p = m + 1$, und es muß mittels der verallgemeinerten Greenschen Funktion $G(\tau, \tau_*)$ zur nichtlinearen Integrodifferentialgleichung

$$
(29)\qquad \zeta(\tau) = \xi\cos p\tau + \eta\sin p\tau + \int_0^{2\pi} G(\tau, \tau_*)\Psi_* d\tau_*
$$

mit

$$
(30)\qquad \xi = \tfrac{1}{\pi}\int_0^{2\pi}\zeta\cos p\tau\ d\tau, \qquad \eta = \tfrac{1}{\pi}\int_0^{2\pi}\zeta\sin p\tau\, d\tau
$$

übergegangen werden.

6. Sind μ, κ und auch ξ, η beliebig vorgegeben in einem gewissen vierdimensionalen Interval Q um den Nullpunkt $\mu = \kappa = \xi = \eta = 0$, so ergibt sich durch sukzessive Approximationen genau eine Lösung $\zeta = \zeta(\tau; \mu, \kappa, \xi, \eta)$

von (29), die sich in eine Potenzreihe nach κ, μ, ξ, η entwickeln läßt. Damit die gewonnene Funktion ζ, die in τ mod 2π periodisch ist und die vorgegebenen Werte ξ, η als Parameter im Sinne von (30) hat, aber auch eine Lösung der ursprünglichen Lagrangeschen Differentialgleichung (28) ist, muß ζ noch die sog. Verzweigungsgleichungen

$$(31)\quad R \equiv (a/m)^2 \frac{1}{\pi}\int_0^{2\pi} \Psi \cos p\tau \, d\tau = 0, \quad S \equiv (a/m)^2 \frac{1}{\pi}\int_0^{2\pi} \Psi \sin p\tau \, d\tau = 0$$

befriedigen.

Schreiben wir in ζ die Glieder erster Ordnung, die aus der ersten Approximation entnommen werden können, aus, so kommt

$$(32)\quad \zeta = 2a^{-2}m^2(m+1)^{-1}\kappa + \beta\mu + \xi\cos(m+1)\tau + \eta\sin(m+1)\tau + \cdots;$$

der Koeffizient β von μ ist für das folgende belanglos.

Zuerst berechnen wir

$$\begin{aligned}(33)\quad R &= \{[12B_{0030} + 4(m+1)^2B_{0012}](m+1)^{-1} \\ &\quad + (a/m)^2[2B_{1020} + 2(m+1)^2B_{1002}]\}\kappa\xi + 2B^{(p)}\mu + R^{(1)} \\ &= \{-12m(m+1) + (-4+1)(m+1)^2 + 3(m+1)(3m+1)\}\kappa\xi \\ &\quad + 2B^{(p)}\mu + R^{(1)},\end{aligned}$$

also

$$(34)\qquad R = -6m(m+1)\kappa\xi + 2B^{(m+1)}\mu + R^{(1)},$$

wo

$$(35)\qquad 2B^{(m+1)} = (a/m)^2 \frac{1}{\pi}\int_0^{2\pi} B_{0110}\cos(m+1)\tau \, d\tau$$

und $R^{(1)}$ sowie nachher $S^{(1)}$ Glieder zweiter Ordnung mit dem Faktor μ sowie Glieder dritter und höherer Ordnung enthält.

Die analoge Rechnung gibt

$$(36)\qquad S = -6m(m+1)\kappa\eta + S^{(1)};$$

der Koeffizient von μ verschwindet hier offenbar aus Symmetriegründen.

7. Vor allem benötigen wir das Nichtverschwinden des Koeffizienten von μ, $2B^{(p)} = am^{-2}\cdot 2A^{(p)}$ in der Bezeichnung meiner früheren Arbeit *b*), in der ich den diesbezüglichen Nachweis allgemein erbracht habe.

Man führt dazu die Laplaceschen Koeffizienten A_s^p ein, die durch die Fourierentwicklung

$$(37)\qquad (1 - 2\alpha\cos\tau + \alpha^2)^{-s} = A_s^0 + 2\sum_{p=1}^{\infty} A_s^p \cos p\tau, \quad A_s^{-p} = A_s^p$$
$$(0 < \alpha < 1,\ s \text{ reell})$$

definiert sind. Sie können bekanntlich mit Hilfe der Gaußschen hypergeometrischen Reihe dargestellt werden

$$(38)\quad A_s^{\,p} = A_s^{\,|p|} = \frac{\Gamma(s+|p|)}{\Gamma(s)\Gamma(|p|+1)}\,\alpha^{|p|}F(s, s+|p|, |p|+1, \alpha^2).$$

Für uns ist vor allem wichtig, daß für $0 < s$ sämtliche Glieder der Reihe positiv sind.

Wir setzen jetzt $s = \frac{1}{2}$ und unterscheiden:

1. Fall eines direkten unteren Planeten ($\begin{smallmatrix} p=2,3,\cdots \\ m=1,2,\cdots \end{smallmatrix}$). Hier gilt mit $\alpha = a$

$$(39)\quad 2B^{(p)} = 4pA_s^{\,p} - 2\alpha^2 A^p{}_{s+1} + \alpha(A^{p+1}_{s+1} + A^{p-1}_{s+1}).$$

Die $A_s^{\,p}$ genügen nun bekanntlich den Identitäten

$$(40)\quad \begin{aligned} 0 &= A_s^{\,p} - (1+\alpha^2)A^p{}_{s+1} + \alpha(A^{p+1}_{s+1} + A^{p-1}_{s+1}),\\ \frac{\alpha}{s}\frac{d}{d\alpha}A_s^{\,p} &= -A_s^{\,p} + (1-\alpha^2)A^p{}_{s+1}. \end{aligned}$$

Aus (39), $(40)_1$ ergibt sich durch Subtraktion

$$(41)\quad 2B^{(p)} = (4p-1)A^p{}_{1/2} + (1-\alpha^2)A^p{}_{3/2},\quad B^{(p)} = 2pA^p{}_{1/2} + \alpha\frac{d}{d\alpha}A^p{}_{1/2} > 0.$$

2. Fall eines direkten oberen Planeten ($\begin{smallmatrix} p=-1,-2,\cdots \\ m=-2,-3,\cdots \end{smallmatrix}$). Dann ist mit $\alpha = 1/a$

$$(42)\quad 2B^{(p)} = 4p\alpha A_s^{\,p} - 2\alpha A^p{}_{s+1} + \alpha^2(A^{p+1}_{s+1} + A^{p+1}_{s+1});$$

dies mit a multipliziert und davon $(40)_1$ subtrahiert, gibt

$$(43)\quad \begin{aligned} 2aB^{(p)} &= (4p-1)A^p{}_{1/2} - (1-\alpha^2)A^p{}_{3/2},\\ aB^{(p)} &= (2p-1)A^p{}_{1/2} - \alpha\frac{d}{d\alpha}A^p{}_{1/2} < 0. \end{aligned}$$

(In der Arbeit *b*) hiess übrigens in diesem 2. Fall $aB^{(p)}$ einfach $B^{(p)}$).

8. Für $\mu = 0$ haben die Keplerschen Kreise verschwindende Parameter ξ, η. Die eigentlichen Keplerellipsen müssen die feste grosse Halbachse a, gegeben durch $(20)_2$ mit $m = 1, \pm 2, \cdots$, haben, um im synodischen Achsenkreuz x, y die Winkel-Periode 2π zu haben. Was die übrigen Elemente betrifft, so nehmen wir als Epoche $t = 0$ den Zeitmoment, wo die Bahn von P_0 die positive Syzygienachse kreuzt. Von dieser, als X_0-Achse siderisch fixiert, zählen wir (alles siderisch) die mittlere Länge $\lambda = nt$, die wahre Länge v und die Perihellänge ϖ; die mittlere Länge der Epoche ist $\epsilon = 0$. Die Exzentrizität e sowie $|\varpi|$ oder $|\varpi - \pi|$ seien klein vorausgesetzt. Sei noch die exzentrische Anomalie u, die Flächenkonstante $c = na^2(1-e^2)^{\frac{1}{2}}$, so gelten die Beziehungen

(44) $\lambda - \varpi = u - e \sin u,\ r = a(1 - e\cos u),\ dv = cr^{-2}dt = (c/n)r^{-2}d\lambda.$

Somit gelten für Radiusvektor und wahre Länge die Potenzreihen nach e:

(45) $r = a(1 - e\cos(\lambda - \varpi) + \cdots),\quad v = \lambda + 2e\sin(\lambda - \varpi) + \cdots.$

Die Ellipsen, für die $|\varpi|$ bezw. $|\varpi - \pi|$ klein sind, bilden eine zusammenhängende stetige Schar, die man erhält, wenn man $|\varpi|$ klein und $-1 < e < +1$ nimmt.

Führen wir jetzt den von der Syzygienachse gezählten Polarwinkel ein:

(46) $\tau = v - t = \lambda(1 - n^{-1}) + 2e\sin(\lambda - \varpi) + \cdots,$

(47) $p\tau = (m+1)\tau = \lambda + 2pe\sin(\lambda - \varpi) + \cdots,$

so ist umgekehrt

(48) $\lambda = p\tau - 2pe\sin(p\tau - \varpi) + \cdots,$

(49) $\cos(\lambda - \varpi) = \cos(p\tau - \varpi) + 2pe\sin^2(p\tau - \varpi) + \cdots.$

Nach $(45)_1$ wird jetzt für eine Keplerellipse die durch (14) definierte Grösse $\zeta = \zeta(\tau|e, \varpi)$ eine Potenzreihe in e, ϖ deren Koeffizienten (sogar schon mod $2\pi/p$) periodische Funktionen von τ sind. Unsere Parameter werden

(50) $\xi = -e + \cdots,\qquad \eta = -e\varpi + \cdots.$

Es ist daher auch $\zeta = \zeta(\tau|\xi, \varpi)$ eine Potenzreihe in ξ, ϖ.

Für die betrachteten Keplerellipsen ist die Jacobische Konstante die Differenz von absoluter Energiekonstante h und Flächenkonstante c

(51) $k = h - c,\ h = h_0 = -1/(2a),\ c = c_0(1 - e^2)^{\frac{1}{2}},\ c_0 = a^{\frac{1}{2}} = n^{-1/3}.$

Es ist daher

(52) $\kappa = -(c - c_0) = -c_0[(1 - e^2)^{\frac{1}{2}} - 1] = \frac{1}{2}c_0\xi^2 + \cdots = \kappa(\xi, \varpi)$

eine Potenzreihe in ξ, ϖ, die der Gleichung

(53) $P(\kappa, \xi, \varpi) \equiv \kappa - \frac{1}{2}c_0\xi^2 + \cdots = 0$

genügt.

9. Bei der Diskussion der Verzweigungsgleichungen (31) muß ich mich—jedenfalls im Moment—auf die Bahnen beschränken, die zur Syzygienachse symmetrisch sind: $\zeta(-\tau) = \zeta(\tau)$. (Für $\mu = 0$ lassen wir in den Potenzreihen ζ, κ, P der vorigen Nr. das Argument $\varpi = 0$ einfach weg). Es ist dann von vornherein $\eta = 0$ zu setzen, und dies ist auch hinreichend dafür, daß die Lösung der Integrodifferentialgleichung (29) die Symmetrie bzgl. der Sygy-

gienachse hat. Die zweite Verzweigungsgleichung $S = 0$ ist dann identisch in μ, κ, ξ erfüllt; es muß also S den Parameter η als Faktor enthalten.

Wir haben nur noch die Bedingung $R = 0$ für die Parameter μ, κ, ξ zu studieren. Da $B^{(p)} \neq 0$ ist, bestimmt sich der Massenprozentsatz als Potenzreihe in κ, ξ

$$(54) \qquad \mu = f(\kappa, \xi) = b\kappa\xi + \cdots, \qquad b = 3m(m+1)/B^{(m+1)} \neq 0.$$

Nun beachte man: die Potenzreihe $\zeta(\tau|\xi)$ in ξ zusammen mit der Potenzreihe (52) $\kappa = \kappa(\xi)$ und $\mu = 0$ in den Ausdruck R eingesetzt, gibt eine Potenzreihe in ξ, die identisch verschwindet. Auf der Lösungsfläche (54) im Raum der Koordinaten κ, ξ, μ liegen also jedenfalls die Punkte der parabelartigen Kurve (53) $P(\kappa, \xi) = 0$ in der Ebene $\mu = 0$, ferner, wie man ebenso sieht, die Punkte der Geraden $\xi = 0$, $\mu = 0$. Auf jeder der beiden im Ursprung irreduziblen algebroiden Kurven muß also sicher $f(\kappa, \xi) = 0$ sein. Daher gilt mit einer Potenzreihe $E(\kappa, \xi)$ die Teilbarkeit [11]

$$(55) \qquad f(\kappa, \xi) = \xi P \cdot E = \xi(\kappa - \tfrac{1}{2}c_0\xi^2 + \cdots)E(\kappa, \xi), \quad E(0, 0) = b \neq 0.$$

Der Punkt $\kappa = \xi = 0$ ist ein Sattelpunkt der Lösungsfläche (54). Die Kurven $\mu = \text{const.}$ auf der Lösungsfläche haben in der Nähe von $\kappa = \xi = 0$ hyperbelartigen Charakter.[12]

Betrachtet man (etwa unter der Voraussetzung $m = 1, 2, \cdots$) die zum festen Wert $\kappa_1 < 0$ gehörige "natürliche" Lösungsschar Σ_1 (erster Sorte) sowie die zum festen Wert $\kappa_2 > 0$ gehörige natürliche Lösungsschar Σ_2 (zweiter Sorte mit $\xi < 0$)—natürlich nur innerhalb des Quaders Q —, so kann man wohl sagen, daß jeder Lösung von Σ_1 durch Gruppenfortsetzung auf der Lösungsfläche (also bei festem $\mu > 0$) eine Lösung von Σ_2 entspricht. Analog sind die Verhältnisse für positiven Exzentrizitätsparameter ξ, dem bei $\mu = 0$ die Perihellänge $\varpi = \pi$ entspricht. Nur sind jetzt beide $\kappa_1, \kappa_2 > 0$, sie können also auch gleich sein.

Auch die zum festen Wert $\kappa = 0$ gehörige natürliche Schar Σ_0 existiert. Es ist dann

$$(56) \qquad \mu = -\tfrac{1}{2}bc_0\xi^3 + \cdots,$$

also

$$(57) \qquad \xi = (-2b^{-1}c_0^{-1}\mu)^{1/3} + \cdots.$$

Diese vom kritischen Kreis ausgehende natürliche lineare Reihe Σ_0 von

---

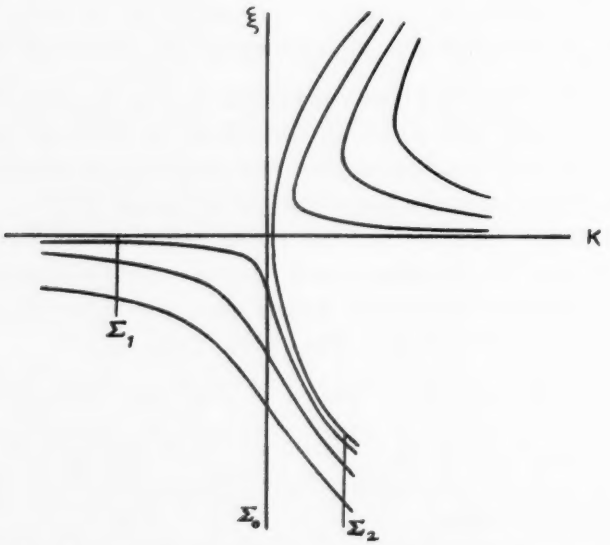
[11] Vgl. Osgood, *Lehrbuch der Funktionentheorie* II_1, Leipzig-Berlin 1924, p. 87, 2. Zusatz. Die entsprechende algebraische Tatsache gründet auf einem von meinem Vater bewiesenen Satz: O. Hölder, "Zum Invariantenbegriff," *Math.-naturw. Mitt. Württ.*, Bd. 1 (1884), S. 60-62.

[12] Vgl. Hadamard, *Cours d'analyse* I, p. 367 f., Paris, 1927.

periodischen Lösungen entgeht der Betrachtungsweise von Martin, der $\mu = x^4$ schreibt und, *loc. cit.*, p. 267,

$$\zeta = O(x^3) = O(\mu^{3/4}) \tag{58}$$

voraussetzt.



10. Auf der Lösungsmannigfaltigkeit liegt, wie noch gezeigt werden soll, die vom kritischen Kreis K^0 ausgehende isoperiodische Schar (der Lösungen von der synodischen zeitlichen Periode $T_0 = 2\pi m$) *in* der Ebene $\mu = 0$, sie besteht also nur aus den Keplerellipsen vom Periodenquotient m.[13]

Für die Lösungen der Integrodifferentialgleichung (29) ist die zeitliche Periode, durch (8) definiert,

$$T = T(\mu, \kappa, \xi) = \int_0^{2\pi} [2(U+k)]^{-\frac{1}{2}} a[\zeta'^2 + (1+\zeta)^2]^{\frac{1}{2}} d\tau. \tag{59}$$

Setzen wir hier (54) zufolge $\mu = b\kappa\xi + \cdots$ ein, so ergibt sich (auf der Lösungsfläche)

$$T = \mathfrak{T}(\kappa, \xi). \tag{60}$$

Es geht nun

$$\frac{\partial \mathfrak{T}}{\partial \kappa} = \frac{\partial T}{\partial \kappa} + \frac{\partial T}{\partial \mu}\frac{\partial \mu}{\partial \kappa} \to \left(\frac{\partial T}{\partial \kappa}\right)_0 \text{ für } \kappa, \xi \to 0, \tag{61}$$

und es kann bei der an der Stelle $\mu = \kappa = \xi = 0$ zu nehmenden Ableitung rechts vor der Differentiation $\mu = \xi = 0$ gesetzt werden, d. h. man berechnet den Limes längs der von κ abhängigen Lösungsschar $\mu = \xi = 0$ von (29),

[13] Vgl. Wintner, *loc. cit.*, 1, a) S. 49.

die offenbar die Schar der zu $\mu = 0$ gehörigen Keplerschen Kreise mit den Halbachsen $a(1+\gamma)$ sein muß. In dieser Schar ist

$$\mathbf{T} = 2\pi(a^{-3/2}(1+\gamma)^{-3/2} - 1)^{-1} = \mathbf{T}_0(1 - \tfrac{3}{2}a^{-3/2}(a^{-3/2} - 1)^{-1}\cdot\gamma + \cdots)^{-1} = \mathbf{T}_0(1 + \tfrac{3}{2}(m+1)\gamma + \cdots) \tag{62}$$

und

$$k = -\tfrac{1}{2}a^{-1}(1+\gamma)^{-1} - a^{1/2}(1+\gamma)^{1/2} = k^0 + \tfrac{1}{2}(a^{-1} - a^{1/2})\gamma + \cdots = k^0 + \tfrac{1}{2}a^{1/2}m^{-1}\gamma + \cdots, \tag{63}$$

also für $\kappa = 0$

$$\frac{\partial \mathbf{T}}{\partial \kappa} = 3\mathbf{T}_0(m+1)ma^{-1/2} = 6\pi m^3 a^{-2} \neq 0. \tag{64}$$

Da sich in (61) nunmehr der Limes rechts von Null verschieden ergeben hat, geht durch $\kappa = \xi = 0$ genau ein linearer regulärer Zweig $\kappa = \kappa(\xi)$ der Kurve $\mathfrak{T}(\kappa, \xi) = \mathbf{T}_0$ hindurch. Dieser muß notwendig (52) $\kappa = \frac{1}{2}c_0\xi^2 + \cdots, \mu = 0$ sein.

11. Nur andeuten kann ich die Rechnungen, die auszuführen wären, um in der Nachbarschaft des Keplerkreises K^0 der mittleren Bewegung

$$n = p/q, \quad p, q \text{ ganz}, (p, q) = 1, \quad p - q > 0, \tag{65}$$

dieser $p - q = J$ mal syndosch durchlaufen, die (zur Syzygienachse symmetrischen) periodischen Bahnen zu übersehen, die sich (mindestens) nach dem synodischen Umlauf $2\pi J$ schließen. Es ist dann der Periodenquotient

$$m = q/(p-q) = q/J \quad \text{und} \quad m + 1 = p/J, \qquad ((p, J) = 1). \tag{66}$$

Multiplizieren wir die Differentialgleichung (28) mit J^2 und setzen

$$\tau/J = \delta, \tag{67}$$

so kommt

$$\frac{d^2\zeta}{d\delta^2} + p^2\zeta = \bar{\Psi}, \qquad (\bar{\Psi} = J^2\Psi). \tag{68}$$

Davon sind die im Intervall $0 \leqq \delta \leqq 2\pi$ periodischen Lösungen zu suchen. Das zugehörige Intervall $0 \leqq \tau \leqq 2\pi J$ heißt das primitive Intervall.

Was die Entwicklung der Lösung ζ der zu (68) gehörigen Integrodifferentialgleichung anlangt, so ändert sich, da ja die Differentialgleichung dieselbe geblieben ist, an dem konstanten Glied erster Ordnung gar nichts. Wir haben

$$\zeta = 2a^{-2}m^2(m+1)^{-1}\kappa + \beta\mu + \xi\cos p\delta + \eta\sin p\delta + \cdots = 2a^{-2}m^2(m+1)^{-1}\kappa + \beta\mu + \xi\cos(m+1)\tau + \eta\sin(m+1)\tau + \cdots \tag{69}$$

nach wie vor, wo jetzt

$$(70)\qquad \begin{aligned} \xi &= \pi^{-1}\int_0^{2\pi} \zeta \cos p\delta\, d\delta = (J\pi)^{-1}\int_0^{2\pi J} \zeta \cos\,(m+1)\tau d\tau, \\ \eta &= \pi^{-1}\int_0^{2\pi} \zeta \sin p\delta\, d\delta = (J\pi)^{-1}\int_0^{2\pi J} \zeta \sin\,(m+1)\tau d\tau \end{aligned}$$

und β ein wieder nicht in Betracht kommender Koeffizient ist.

Setzen wir von vornherein Symmetrie bezüglich der Syzygienachse, d. h. $\eta = 0$ voraus, so brauchen wir nur die eine Verzweigungsgleichung

$$(71)\quad \bar{R} = (a/m)^2\pi^{-1}\int_0^{2\pi} \Psi \cos p\delta\, d\delta = (a/m)^2\pi^{-1}J\int_0^{2\pi J} \Psi \cos\,(m+1)\tau d\tau = 0$$

zu betrachten. Darin berechnet sich der Koeffizient von $\kappa\xi$ in derselben Weise wie früher für $J = 1$, nur hat man zum Schluß das Integral

$$(72)\qquad \int_0^{2\pi J} \cos^2(m+1)\tau d\tau = J\int_0^{2\pi} \cos^2 p\delta\, d\delta = J\pi.$$

Ein Glied mit μ allein tritt in (71) für $J > 1$ nicht auf. Der Koeffizient $\bar{B}_{0110} = J^2 B_{0110}$ von μ in $\bar{\Psi}$ läßt sich nämlich in eine Kosinusreihe nach $\tau = J\delta$ entwickeln; ein Vielfaches dieses Winkels kann aber niemals gleich dem Winkel $p\delta$ sein, mit dessen Kosinus $\bar{\Psi}$ in (71) zu falten ist. Denn daß ein Vielfaches $NJ = p$ wird, ist bei der Teilerfremdheit von p und J nur für $J = 1$ möglich. Im ganzen bekommen wir für (71)

$$(73)\qquad \bar{R} \equiv J^2 \cdot 6m(m+1)\kappa\xi + \cdots = \xi(6pq\kappa + \cdots) = 0.$$

Dazu muß man noch beachten, daß $\bar{R} = 0$ gelten muß, sowie $\xi = 0$ ist. Denn für $\xi = 0$ hat die zu $0 \leqq \tau \leqq 2\pi J$ gehörige Integrodifferentialgleichung eine eindeutig bestimmte, in κ, μ reguläre analytische Lösungsschar. Diese muß man bekommen, wenn man von der zu $0 \leqq \tau \leqq 2\pi$ gehörigen Integrodifferentialgleichung ausgeht; deren Lösungen sind ja, da der reguläre Fall vorliegt, ohne weiteres Lösungen der Lagrangeschen Gleichung für die Bahnkurven und haben als Kosinusreihen in τ wieder wegen der Teilerfremdheit von p, J vermöge (70) den Parameter $\xi = 0$, genügen also der zu $0 \leqq \tau \leqq 2\pi J$ gehörigen Integrodifferentialgleichung mit $\xi = 0$.

Die Verzweigungsleichung (73) zeigt, daß die Mannigfaltigkeit der symmetrischen Lösungen zerfällt; auch bei festem μ erhalten wir zwei Lösungsgruppen, entsprechend Bahnen mit der synodischen Windungszahl 1 bezw. J.

Weitere Rechnungen bedürfte die Bestätigung der Vermutung, daß die Zeitspanne $JT_0 = 2\pi mJ = 2\pi q$ selbst als primitive synodische Periode von Bahnen mit $\mu > 0$ nicht vorkommt.

Universität Leipzig.

SYMBOLIC DYNAMICS.

By Marston Morse and Gustav A. Hedlund.

---

1. Introduction. In both its qualitative and quantitative aspects, modern theoretical dynamics is largely concerned with recurrence and transitivity. The qualitative study of dynamics owes its inception to Hill and Poincaré. Its development in the past three decades has been dominated by the work of Birkhoff, to whom belongs the principal credit for the trend of modern research in dynamics. The quantitative (metrical) study of dynamical systems is of more recent origin, but with the establishment of the ergodic theorems of Birkhoff and von Neumann, its development has been rapid.

The methods used in the study of recurrence and transitivity frequently combine classical differential analysis with a more abstract symbolic analysis. This involves a characterization of the ordinary dynamical trajectory by an unending sequence of symbols termed a *symbolic trajectory* such that the properties of recurrence and transitivity of the dynamical trajectory are reflected in analogous properties of its symbolic trajectory. For general research of this character one may refer to papers by Hadamard, Birkhoff and the authors. Papers which are special in that they refer to spaces of constant curvature, but which are significant by virtue of the symbolisms involved have been written by Artin, Koebe, Martin, Myrberg and Nielsen. Martin and Robbins have contributed papers on purely symbolic aspects.

In the general theory as developed up to the present, the dividing line between symbolic and differential methods has not been sharply defined. It is a first aim of the present paper to separate the symbolic aspects from the differential aspects and to develop the symbolic theory on its own account. In this way many interesting and important dynamical questions can be given an algebraic formulation quite apart from the classical theories of differential equations.

A symbolic trajectory T is formed from symbols taken from a finite set of generating symbols subject to rules of admissibility which are dependent on the underlying Poincaré fundamental group. The trajectory T taken with one of its symbols is called a *symbolic element*. Symbolic elements are analogues of line elements on an ordinary trajectory. A metric is assigned to the space of symbolic elements and this space turns out to be perfect, compact, and totally disconnected. The necessary first theorems on symbolic recurrence and transitivity are then rapidly developed in a way suggestive of the differential treatment of Birkhoff.

We come to the more novel developments of the paper. Recall that the recurrency index $R(n)$ of T is a minimum number m such that each block of m successive symbols of T contains every block of n successive symbols of T. If $R(n)$ exists for each n, T is termed recurrent (cf. Morse [2], p. 94). In § 7 we are concerned with conditions on $R(n)$. A typical condition is that $R(n) > 2n$ for every non-periodic recurrent T. In § 8 we make a special study of the Morse recurrent trajectory (Morse [2], p. 95) obtaining $R(n)$ explicitly. This is the first and only known evaluation of $R(n)$ for a non-periodic recurrent trajectory. We give a definition of almost periodicity of symbolic trajectories somewhat similar to the ordinary definition of Besicovitch ([1], p. 77) and show that the Morse trajectory is not almost periodic. Most of the known recurrent trajectories are almost periodic. In § 9 we give general methods for deriving new recurrent trajectories from a given recurrent trajectory. For the Morse trajectory, $R(n)/n$ is shown to have an inferior limit of 5.5 and a superior limit of 10. In § 10 we define a recurrent trajectory for which the inferior limit of $R(n)/n$ is finite while the superior limit is $+\infty$.

We continue with a study of transitivity. An admissible symbolic sequence of the form $a_1a_2a_3\cdots$ is termed a *ray*. We assign the Baire metric (cf. Hausdorff [1], p. 101) to the space of rays. Let $\theta(n)$ be the minimum length of a block which contains a copy of each admissible n-block. A ray R which contains a copy of each admissible block is termed transitive. If R is transitive, let $\phi(n)$ be the minimum length of an initial block of R which contains a copy of each admissible n-block. We generalize a theorem of Martin [1] and include a hitherto unpublished theorem of R. Oldenburger on the impossibility that $\phi(n) = \theta(n)$ for two successive values of n $(n > 1)$. A typical general theorem states that the set of transitive rays for which

$$\liminf_{n\to\infty} \frac{\phi(n)}{\theta(n)} = 1$$

is a residual set of the second category.

Symbolic dynamics as the authors conceive it forms one of the three divisions

(1) representation theory,
(2) symbolic dynamics,
(3) existence of space forms,

of the whole theory. The representation theory is concerned with the conditions on space forms under which trajectories admit a one-to-one symbolic representation in terms of which the recurrence or transitivity of the trajectory can be determined. These conditions will involve the Poincaré fundamental

group of the space and differential conditions such as that of uniform instability (cf. Morse [4], p. 64). In (3) one is concerned with the existence of space forms satisfying the conditions discovered in (1). The questions involved are rather deep extensions of the Hilbert, Koebe theory of spaces of negative curvature (cf. Hilbert [1], and Koebe [1]). A simple typical theorem is that there exists no two-dimensional Riemannian manifold of the topological type of the torus satisfying the condition of uniform geodesic instability. The bearing of such studies on questions of topological and metric transitivity will be made clear in later papers.

2. Symbolic sequences. Having seen the significance of symbolic sequences we shall now develop their theory on an independent basis. We start with a finite set of μ distinct symbols termed *generating symbols.* We suppose that $\mu > 1$. If a_n represents a generating symbol, sequences of the form

$$\cdots a_{-1}\, a_0\, a_1 \cdots, \tag{2.1}$$

$$a_r\, a_{r+1}\, a_{r+2} \cdots, \tag{2.2'}$$

$$\cdots a_{m-2}\, a_{m-1}\, a_m, \tag{2.2''}$$

$$a_{r+1}\, a_{r+2} \cdots a_{r+n}, \tag{2.3}$$

will be termed indexed *sequences* (written I-sequences). More particularly I-sequences of the forms (2.1), (2.2) and (2.3) will be termed *I-trajectories, I-rays,* and *I-blocks* respectively. It will be convenient to term the generating symbol represented by a_n the *value* of a_n. The I-blocks (2.3) will be said to have the *length* n and will frequently be termed n-blocks.

Suppose A and B are I-sequences of symbols a_n and b_m respectively. We shall regard A and B as identical and write $A \equiv B$ if and only if the ranges of the indices n and m are the same and a_n and b_n have the same value. Let r be an integer, positive, negative or zero. Suppose that the ranges of n and m are such that symbols a_n and b_m correspond in a one-to-one manner with a_n corresponding to b_{n+r}. If then a_n and b_{n+r} have the same value we shall say that A and B are *similar* and write $A \sim B$. We shall also write

$$B \equiv D_r A, \qquad A \equiv D_{-r} B.$$

If $r = 0$, $D_r A \equiv A$, but in general $D_r A$ is different from A when $r \neq 0$.

The class of I-sequences similar respectively to an I-trajectory, I-ray, or I-block will be termed a *trajectory, ray,* or *block,* and will be represented by the sequence of values involved without indices. For example if 1 and 2 are generating symbols, the unending sequence

$$\cdots 1\,2\,1\,2\,1\,2 \cdots$$

is a trajectory but not an I-trajectory.

Corresponding to an I-sequence A we shall refer to *I-subblocks* or *I-subrays*. We shall thereby mean subsets of consecutive symbols of A 'indexed as' in A with successive indices differing by unity. Thus if A is of the form

$$a_1\, a_2\, a_3\, a_4,$$

a_2a_3 is an I-subblock, but a_1a_3 is not an I-subblock.

In classical dynamics an element based on a motion T is a point P of T together with the direction tangent to T at P. The symbolic analogue of such an element is defined as follows. Let (a) be an I-trajectory and a_r the r-th symbol of (a). The I-trajectory (a) taken with a_r will be termed an I-element based on (a) at a_r, and will be denoted by $E(r, a)$. We term r the *index* of $E(r, a)$ and a_r the *preferred* symbol. Let $E(s, b)$ be an I-element of index s based on an I-trajectory (b). We shall regard $E(r, a)$ and $E(s, b)$ as identical and write

$$E(r, a) \equiv E(s, b)$$

if and only if $(a) \equiv (b)$ and $r = s$. We shall say that $E(r, a)$ is similar to $E(s, b)$ and write $E(r, a) \sim E(s, b)$ if

$$D_{-r}(a) \equiv D_{-s}(b).$$

The class of I-elements similar to a given I-element $E(r, a)$ will be termed a *symbolic element* E. Two symbolic elements are regarded as identical if the corresponding classes of I-elements are identical. We say that E is based on (a) and represented by $E(r, a)$.

We shall give an illustrative example. Let (a) be an I-trajectory in which a_n equals 1 or 2 according as n is even or odd, and let (b) be an I-trajectory in which $b_{n+1} = a_n$. Evidently

$$E(0, a) \not\equiv E(0, b), \qquad E(0, a) \sim E(1, b).$$

More generally if n is even and m is odd,

$$E(n, a) \sim E(m, b)$$

and these similar elements serve to define a symbolic element E. There are just two different symbolic elements based on (a) or (b).

3. The space of symbolic elements. We shall assign a distance E_1E_2 to any two symbolic elements E_1, E_2. To that end we say that an I-block of the form

$$c_{r-m} \cdot \cdot \cdot c_r \cdot \cdot \cdot c_{r+m}$$

is *centered* at c_r. Let E_1 and E_2 be represented by indexed elements $E(r, a)$ and $E(s, b)$ respectively. If $E_1 \neq E_2$, let m be the length of the longest I-blocks of (a) and (b) which are similar and centered at a_r and b_s respectively. We then set

$$E_1E_2 = E(r, a)E(s, b) = 1/m \qquad (m \neq 0). \tag{3.1}$$

In the special case where $m = 0$ we regard the distance (3.1) as infinite. If $E_1 \equiv E_2$, the distance (3.1) is taken as zero. It is clear that the distance E_1E_2 is independent of the particular I-elements $E(r, a)$ and $E(s, b)$ chosen to represent E_1 and E_2 respectively.

The distances so defined satisfy the usual metric axioms:

$$E_1E_2 = 0 \text{ if and only if } E_1 \equiv E_2, \tag{3.2}$$

$$E_1E_2 = E_2E_1, \tag{3.3}$$

$$E_1E_3 \leqq E_1E_2 + E_2E_3. \tag{3.4}$$

Relation (3.4) follows from the stronger relation

$$E_1E_3 \leqq \max(E_1E_2, E_2E_3). \tag{3.5}$$

In case any one of the distances involved in (3.5) is 0 or ∞, (3.5) holds trivially as is readily seen. Excepting these cases, let h, k, and m be the reciprocals of E_1E_3, E_1E_2 and E_2E_3 respectively. Consider the case $k \geqq m$. Evidently then $h \geqq m$ so that $h \geqq \min(k, m)$. The latter relation yields the condition

$$1/h \leqq \max(1/k, 1/m)$$

from which (3.5) follows. The case $m \geqq k$ leads to the same result.

Recall that a metric space M is compact if each infinite sequence of points of M contains at least one subsequence which converges to a point of M. With this understood we shall prove the following.

THEOREM 3.1. *The space M of all symbolic elements is compact.*

Let E_n be the n-th element in an arbitrary sequence of symbolic elements. Suppose E_n is based on the I-trajectory (a^n). Without loss of generality we can suppose that the preferred symbol in (a^n) has the index 0. We shall define a doubly infinite matrix of I-trajectories

$$T_{ij} \qquad (i, j = 1, 2, \cdots) \tag{3.6}$$

as follows. Since there is at most a finite number of generating symbols there exists an infinite subset of the symbols

$$a_0{}^n \qquad (n = 1, 2, \cdots)$$

which have the same "value" c_0. Let the I-trajectories T_{1j} be a subsequence of the I-trajectories (a^n) in which a_0^n has the value c_0. Proceeding inductively we suppose that the first $k-1$ rows of the matrix (3. 6) have been determined. The k-th row shall then be a subsequence T_{ki} of the $(k-1)$-st row such that the symbols of T_{ki} of index $k-1$ have a common value c_{k-1}, while the symbols of index $1-k$ have a common value c_{1-k}.

Let (c) be the I-trajectory

$$\cdots c_{-1}\, c_0\, c_1 \cdots .$$

Then (c) has an I-block of length $2k-1$, with center at c_0, in common with T_{kk}. Let $\mathcal{E}_k$ be the element of index 0 based on T_{kk}, and let $\mathcal{E}$ be the element of index 0 based on (c). Then

$$\mathcal{E}\mathcal{E}_k \leqq \frac{1}{2k-1},$$

so that the elements $\mathcal{E}_k$ converge to $\mathcal{E}$. Moreover the elements $\mathcal{E}_k$ are a subsequence of the given elements $\mathcal{E}_n$, and the proof of the theorem is complete.

A metric space is termed *totally disconnected* if its connected subsets all reduce to points. Cf. Hausdorff [1], p. 152. Recall also that if C is a connected set and d is a distance assumed by a pair of points of C, there exists a pair of points of C which assumes an arbitrary distance between 0 and d. In the case of the space M of symbolic elements, the distances take on only the discrete values $1/n$, so that the space M is totally disconnected.

Finally, the space M is perfect. We already know that M is compact. To prove that M is perfect it is necessary to show that in every neighborhood of an indexed element $E(0, a)$ there is an I-element $E(0, b)$ not similar to $E(0, a)$. To that end we take (b) so that it has a prescribed block A of (a) centered at a_0, in common with (a), but that all longer blocks of (a) with centers at a_0 differ from corresponding blocks of (b). The element $E(0, b)$ will not be similar to $E(0, a)$ but will have an arbitrarily small distance from $E(0, a)$ if the block A is sufficiently long. We accordingly have the following theorem.

THEOREM 3. 2. *The space of all symbolic elements is compact, perfect and totally disconnected.*

By virtue of a well known theorem in point set theory (cf. Hausdorff [1], p. 160), Theorem 3. 2 has the following corollary.

COROLLARY. *The space of all symbolic elements is the homeomorph of the Cantor, perfect, nowhere dense, linear set.*

4. The space of rays and the space of *I*-trajectories. Let R_1 and R_2 be rays with indexed representations of the form

$$a_r\, a_{r+1} \cdots .$$

If $R_1 \neq R_2$, the rays R_1 and R_2 have common initial blocks of finite maximum length m. With Baire (cf. Hausdorff, p. 101) we then set

$$R_1 R_2 = \frac{1}{m+1}.$$

If $R_1 \equiv R_2$, we set $R_1 R_2 = 0$. The distance so defined satisfies the usual metric axioms including a stronger triangle relation similar to (3.5). One readily sees that the space of rays is compact, perfect and totally disconnected.

This distance function in the space of rays is similar to that previously defined in the space of elements. We define a distance between *I*-trajectories which is based on an entirely different concept and is analogous to the Besicovitch metric of the theory of almost periodic functions (cf. Besicovitch [1], p. 73).

Let (a) and (b) be *I*-trajectories. Let $|\, a_i - b_i \,|$ be 0 if a_i and b_i have the same symbolic value, and 1 if this is not the case. We then set

$$[a, b] = \limsup_{n\to\infty} \frac{1}{2n+1} \sum_{-n}^{n} |\, a_i - b_i \,|.$$

It is easily shown that this distance function satisfies the usual metric axioms with the exception that $[a, b] = 0$ does not imply $(a) \equiv (b)$. A simple example proving this last statement is obtained by letting all the symbols of (a) be 1, and all those of (b) be 1 with the exception of the symbol of index 0 which shall be 2.

If $(a') \sim (a)$ and $(b') \sim (b)$, it will not in general be true that $[a', b'] = [a, b]$, but corresponding to (a') there always exists a $(b') \sim (b)$ such that the equality does hold. Recalling the notation of § 2, let $D_r(a)$ denote the *I*-trajectory (a') in which $a'_{i+r} = a_i$. Thus $D_r(a) \sim (a)$. Furthermore, if $(a') \sim (a)$, there exists an r such that $D_r(a) \equiv (a')$. We shall establish the following lemma.

LEMMA 4.1. *If (a) and (b) are two I-trajectories and r is integral,*

$$[D_r(a), D_r(b)] = [a, b]. \tag{4.1}$$

Since $D_r(D_{-r}(a)) = (a)$, it is sufficient to prove (4.1) under the assumption that $r > 0$. Denoting $D_r(b)$ by (b') we have

$$\frac{1}{2n+1}\sum_{-n}^{n}|a'_i-b'_i| = \frac{1}{2n+1}\sum_{-n-r}^{n-r}|a_i-b_i|$$
$$= \frac{1}{2n+1}\sum_{-n}^{+n}|a_i-b_i| + \frac{1}{2n+1}\sum_{-n-r}^{-n-1}|a_i-b_i| - \frac{1}{2n+1}\sum_{n-r+1}^{n}|a_i-b_i|.$$

Since $0 \leqq |a_i - b_i| \leqq 1$, the last two terms approach 0 as n becomes infinite. Thus

$$\limsup_{n\to\infty}\frac{1}{2n+1}\sum_{-n}^{n}|a'_i-b'_i| = \limsup_{n\to\infty}\frac{1}{2n+1}\sum_{-n}^{n}|a_i-b_i|,$$

which is the stated result.

Let T be a trajectory and let (a) be an I-trajectory representing T. Given $\epsilon > 0$, T will be said to admit the *ϵ-translation* r if $[D_r(a), a] < \epsilon$. From (4. 1), this definition is independent of the I-trajectory chosen to represent T. Any trajectory admits the ϵ-translation 0, and this will be termed the *trivial* ϵ-translation.

As in the theory of almost periodic functions, a set of integers $\cdots < r_{-1} < r_0 < r_1 < r_2 < \cdots$ is termed *relatively dense* if the numbers $|r_i - r_{i-1}|$, $i = 0, \pm 1, \cdots$, form a bounded set. We define the trajectory T as *almost periodic* if given $\epsilon > 0$, T admits a relatively dense set of ϵ-translations. A periodic trajectory is evidently almost periodic. We shall presently define recurrent trajectories and shall use the notion of almost periodicity in the classification of such trajectories. Most of the known examples of recurrent trajectories turn out to be almost periodic, but the Morse recurrent trajectory of § 8 is not almost periodic, as we shall see.

5. Restricted sequences. In the applications to dynamical theory the sequences admitted are not formed freely from the generating symbols, but are subject to certain limitations. The simplest of these restrictions may be called the *restriction of inverses*. In this case the generating symbols form a set of the form

$$a_1, \cdots, a_p, a_1^{-1}, \cdots, a_p^{-1}, \qquad (p > 1)$$

in which a_i^{-1} is termed the inverse of a_i. In this case a sequence is inadmissible if a symbol and its successor are inverses. In other cases a circular order C of the generating symbols may appear and an integer r, with the restriction that p-blocks in the order C with $p > r$ are to be excluded. Another type of restriction prohibits successive blocks of certain definite forms.

In general problems restrictions of all of these types will appear. What these restrictions are in any problem depends upon the underlying group theory and topology. A detailed study of these restrictions will appear in a later

paper. In the present paper we are fortunately able to abstract the essentials of these conditions as follows.

Certain blocks of the generating symbols will be distinguished and termed admissible. A ray or trajectory all of whose subblocks are admissible will be termed admissible as well as any element based on an admissible trajectory. The conditions of admissibility on blocks are as follows.

(α). *A block is admissible if and only if all its subblocks are admissible.*

(β). *Corresponding to any admissible block A there exist at least two generating symbols e, dependent at most upon the last two symbols of A and such that Ae is admissible. A similar condition holds replacing last by first and Ae by eA.*

(γ). *Corresponding to any two admissible blocks A and B there exists at least one generating symbol e dependent at most upon the last two symbols of A and the first two symbols of B and such that AeB is admissible.*

(δ). *There exist at least two generating symbols α and β such that all blocks formed from α and β alone are admissible.*

These conditions are satisfied when there are at least two generating symbols and their inverses and the only restriction is the restriction of inverses. They are also satisfied by the well known Nielsen symbolism (cf. Nielsen [2], pp. 211-217) for surfaces of genus $p > 1$. Various symbolisms developed by the authors satisfy these conditions, including one which replaces the Nielsen symbolism to advantage.

We state the following theorem.

THEOREM 5.1. *The set of admissible elements form a compact, perfect, totally disconnected subset Ω of the space M of all elements.*

That Ω is compact may be seen as follows. Let E_n be a sequence of admissible elements based respectively on trajectories t_n. Let E be a limit element of the sequence E_n. Suppose that E is based on the trajectory t. Each block in t must appear in some trajectory t_n and accordingly be admissible. Hence t and E are admissible, and Ω is compact.

To show that Ω is perfect let E be an arbitrary admissible element. Represent E by an I-element $E(0, a)$. Let A be a $(2n+1)$-block of (a) centered at a_0. By virtue of admissibility condition (β), the block A can be extended to form an admissible I-trajectory of the form

$$\cdots e_{-n-2}e_{-n-1}Ae_{n+1}e_{n+2}\cdots, \tag{5.1}$$

where the symbols e_i and e_{-i}, $i > n$, can be successively chosen in two different

ways. Let E' be an element of index 0 based on (5.1). With the freedom of choice of the symbols e_i and e_{-i} it appears that the elements E' include elements different from E. But these elements also include elements arbitrarily near E since A is arbitrarily long.

The set Ω is accordingly perfect. It is totally disconnected since the space M is totally disconnected, and the proof of the theorem is complete.

A theorem similar to the above is true of the set of admissible rays, taking the metric of rays into account. Unless otherwise stated, the only trajectories and rays which will be used will be those which are admissible in the sense of this section.

6. Limit trajectories and minimal trajectories. Let t be a trajectory represented by an I-trajectory (a). If there exists a positive integer r such that

$$a_{n+r} = a_n \qquad (n = 0, \pm 1, \pm 2, \cdots),$$

t will be said to have the *period* r. A trajectory of period r is determined by each of its r-blocks. There are accordingly at most countably infinite many distinct periodic trajectories.

Let Σ be an arbitrary set of admissible trajectories. A trajectory t will be termed a *limit* trajectory of Σ if the elements based on trajectories of Σ include among their limit elements at least one element based on t. The set Σ can consist of a single trajectory, for example of the trajectory t given by

$$\cdots 2\,2\,2\,0\,2\,2\,0\,2\,0\,2\,2\,2\,0\,2\,2\,2 \cdots. \tag{6.1}$$

The trajectory $\cdots 2\,2\,2 \cdots$ is a limit trajectory of the trajectory (6.1).

A periodic trajectory has no limit trajectory because it bears at most a finite number of elements. Let t be a non-periodic trajectory, and (a) an I-trajectory representing t. Any two I-elements with indices n and r respectively based on (a), with $n > r$, are dissimilar. For a similarity relation between these two elements would imply that (a) had the period $n - r$, contrary to hypothesis. The elements based on t with positive indices must then have at least one limit element, and we term such elements *positive limit elements* of t. *Negative limit elements* are similarly defined.

A trajectory can have more than one distinct limit trajectory. Letting a^n stand for a block $a \cdots a$ of length n we see that a trajectory of the form

$$\cdots \beta^4\alpha^3\beta^2\alpha\beta^2\alpha^3\beta^4 \cdots \tag{6.2}$$

has the trajectories $\cdots \alpha\alpha\alpha \cdots$ and $\cdots \beta\beta\beta \cdots$ as limit trajectories. More generally let the set of admissible blocks be enumerated in the form

$$A_0, A_1, \cdots.$$

If generating symbols e_n and b_n are successively chosen in accordance with block condition (γ) the trajectory

$$\cdots A_2 b_2 A_1 b_1 A_0 e_1 A_1 e_2 A_2 \cdots \tag{6.3}$$

will be admissible. It will have every admissible trajectory as a limit trajectory. The elements based on the trajectory (6.3) are everywhere dense in the set Ω of admissible elements. Such a trajectory is termed *transitive*.

In the remainder of this section we shall be concerned with a fundamental type of limit trajectories termed *minimal* trajectories. The dynamical counterpart of such trajectories was first introduced by Birkhoff [1]. Birkhoff also developed their general theory. In this connection the symbolic method was first used by Morse ([1], [2]) to establish the existence of non-periodic recurrent motions. Theorems 6.1 and 6.3 of this section and Theorem 7.2 of the following section are the abstract symbolic analogues of dynamical theorems of Birkhoff. The proofs presented here are independent of the theory of differential equations.

Let s be an admissible trajectory and n an arbitrary positive integer. The number of different n-blocks of s is finite. It will be denoted by $P_s(n)$ and termed the n-th *permutation number* of s. If t is a limit trajectory of s,

$$P_s(n) \geqq P_t(n). \tag{6.4}$$

A trajectory s such that the equality holds in (6.4) for every n and every limit trajectory t of s will be termed *minimal*.

We understand that periodic trajectories are minimal. They satisfy the above definition vacuously since they have no limit trajectories. We continue with a proof of the following theorem.

THEOREM 6.1. *Among the positive (or negative) limit trajectories of an admissible non-periodic trajectory T there is at least one admissible minimal trajectory.*

We shall give the proof for the case of positive limit trajectories. To that end we shall define certain symbols inductively as follows:

$H_0 =$ the set of positive limit trajectories of T.
$p_1 =$ the minimum of $P_t(1)$ for t in H_0.
$H_1 =$ the subset of H_0 for which $P_t(1) = p_1$.
. . . . . .
$p_n =$ the minimum of $P_t(n)$ for t in H_{n-1}.
$H_n =$ the subset of H_{n-1} for which $P_t(n) = p_n$.

The set H_0 is not empty since T is non-periodic. Hence H_n is not empty. Moreover

$$H_0 \supset H_1 \supset H_2 \cdots.$$

We continue with a proof of the following statement.

(α). *The set H_n is closed in the sense that it includes its limit trajectories.*

That H_0 is closed follows from the definition of limit trajectories. Proceeding inductively we shall assume that H_{n-1} is closed and prove that H_n is closed. In case the set H_n has no limit trajectories, H_n is certainly closed. Suppose then that H_n has a limit trajectory t. To show that t is in H_n, recall that $H_{n-1} \supset H_n$ and that t is accordingly a limit trajectory of H_{n-1}. Since H_{n-1} is assumed to be closed t is in H_{n-1}. By virtue of the definition of p_n,

$$P_t(n) \geqq p_n. \tag{6.5}$$

On the other hand, t is a limit trajectory of trajectories of H_n, and the n-th permutation numbers of trajectories of H_n equal p_n. Hence

$$P_t(n) \leqq p_n. \tag{6.6}$$

Hence the equality prevails in (6.5) and (6.6), and t is in H_n. Thus H_n is closed.

The sets H_n form a decreasing sequence of closed sets. It follows from the compactness of the space of admissible elements that the intersection H of the sets H_n is closed and non-void. Let s be a trajectory of H. If s is periodic, s is minimal and the proof is complete. If s is not periodic, let t be a limit trajectory of s. Since t and s are both in H_n for every n, t and s have the same n-th permutation number for every n. Hence s is minimal, and the proof is complete.

THEOREM 6.2. *If s is a minimal trajectory and t a limit trajectory of s, then t is a minimal trajectory and s a limit trajectory of t.*

Every n-block of s is in t since s is minimal, so that s is a limit trajectory of t. To show that t is minimal, let r be one of its limit trajectories. Then r is also a limit trajectory of s. Since s is minimal, the n-th permutation numbers of t and r equal that of s, and so equal each other. Hence t is minimal, and the proof is complete.

The set of limit trajectories of a trajectory t will be termed the *derived set* of t. If s is a non-periodic minimal trajectory, each member of its derived set S has s as a limit trajectory, and hence has the same derived set. That

this property is characteristic of non-periodic minimal trajectories is implied by the following theorem.

THEOREM 6.3. *A set S of trajectories which is the derived set of each of its members consists of non-periodic minimal trajectories.*

Since the derived set of a periodic trajectory is empty, no trajectory of the set S can be periodic.

If s is a trajectory of S and t is a limit trajectory of s,

$$P_s(n) \geqq P_t(n). \tag{6.7}$$

But t is in S and thus s is a limit trajectory of t, so that the equality sign must hold in (6.7). Hence s is minimal and the proof is complete.

The elements based on trajectories of S form a compact, perfect, totally disconnected set H. This is behind an earlier result of Morse that there exist discontinuous recurrent motions. Cf. Morse [2], p. 99. The set H is the homeomorph of the Cantor, perfect, nowhere dense, linear set. In particular H has the power $\aleph$ of the continuum. Since there is at most a countable infinity $\aleph_0$ of elements based on each trajectory of S we can conclude that the trajectories of S have a power a such that $\aleph_0 a = \aleph$. But $\aleph_0 a = a$ (cf. Hausdorff, pp. 30-31) so that $a = \aleph$. The set S thus has the power of the continuum.

We shall refer to the following definition. A set of trajectories which is the derived set of each of its members or which reduces to a single periodic trajectory is termed a *minimal set* of trajectories.

7. Recurrent trajectories. Let h be an admissible sequence, finite or infinite, and n a positive integer. A least integer m (if any exists) such that each m-block of h contains every n-block of h will be called the n-th *recurrency index* $R_h(n)$ of h. If h is a ray or a trajectory and $R_h(n)$ exists for each n, h will be termed *recurrent* and $R_h(n)$ the *recurrency function* of h.

We begin with the following theorem.

THEOREM 7.1. *If s is a recurrent trajectory and t is a limit trajectory of s, then t is recurrent and s is a limit trajectory of t. Moreover*

$$R_s(n) = R_t(n).$$

Each block of t appears in s since t is a limit trajectory of s. Let A be an arbitrary n-block of s. A copy of A is found in each block of s of length $R_s(n)$, and hence in each block of t of length $R_s(n)$. Hence $R_t(n)$ exists with

$$R_t(n) \leqq R_s(n). \tag{7.1}$$

Since each block A of s appears in t, s is a limit trajectory of t. The rôles of s and t can then be interchanged so that the equality prevails in (7.1). The proof of the theorem is complete.

The following theorem is the symbolic analogue of a theorem of Birkhoff. Cf. Birkhoff [2], p. 199.

THEOREM 7.2. *A necessary and sufficient condition that a trajectory s be recurrent is that it be minimal.*

We begin by assuming that s is minimal but not recurrent, and seek a contradiction. If s is not recurrent, there exists a block U of s with the following property. There exists a sequence A_n of blocks of s whose lengths become infinite with n and which do not contain U. Without loss of generality we can suppose that the length of A_n is odd. Let E_n then be the element based on s with preferred symbol at the center of A_n. Since the space of admissible elements is compact there exists a subsequence of the elements E_n which converges to an admissible element E. Let t be the trajectory on which E is based. It is clear that t does not contain the block U. Hence $t \neq s$. It follows that t is a limit trajectory of s. Since s is minimal, t and s must then contain the same blocks. From this contradiction we infer that s is recurrent.

It remains to show that s is minimal if recurrent. If s is periodic, s is minimal. Assume then that s is not periodic and let S be the derived set of s. It follows from Theorem 7.1 that S is the derived set of each of its members. We conclude from Theorem 6.3 that s is minimal, and the proof is complete.

It is natural to ask what limitations on recurrency indices $R(n)$ and permutation indices $P(n)$ are imposed by conditions of periodicity or recurrency. We give a partial answer in the following lemmas and theorems.

LEMMA 7.1. *If $P(n)$ is the number of different n-blocks in a trajectory t, $P(n+1) \geqq P(n)$.*

Let the different n-blocks in t be enumerated in the form

$$B_i \qquad [i = 1, 2, \cdots, P(n)], \tag{7.2}$$

and let e_i be the successor of B_i in some position in t. The $(n+1)$-blocks $B_i e_i$ are all different, and the lemma follows immediately.

If (a) is an indexed trajectory, a block B of (a) whose first and last symbols are a_r and a_m will be denoted by $[r, m]$. We shall term r the *first index* of B and m the *final index* of B.

THEOREM 7.3. *A trajectory t for which $P(m+1) = P(m)$ for some m has a period $\omega \leqq P(m)$.*

Let (a) be an indexed representation of t. There are $P(m)$ dissimilar m-blocks B_i in (a). The elements immediately following copies of B_i in (a) must always be the same. Otherwise there would be at least $P(m)+1$ dissimilar $(m+1)$-blocks. Similarly the elements immediately preceding copies of B_i must always be the same.

Consider the blocks $B=[i+1, i+\mathrm{m}]$. Since there are just $P(m)$ dissimilar m-blocks, for some i there must exist a similarity relation of the form

$$[i+1, i+m] \sim [i+1+\omega, i+m+\omega], \tag{7.3}$$

in which $0<\omega \leqq P(m)$. We shall show that (a) has a period ω. We begin with an inductive proof of the relations,

$$a_{i+r}=a_{i+r+\omega} \qquad (r=1,2,\cdots). \tag{7.4}$$

That (7.4) holds for $r=1,2,\cdots,m$ follows from (7.3). Assuming that (7.4) holds for $r=1,2,\cdots,n \geqq m$, we shall prove that (7.4) holds for $r=n+1$. The m-blocks with final indices $i+n$ and $i+n+\omega$ respectively are similar by virtue of our inductive hypothesis. It then follows from the principle stated in the first paragraph of this proof that (7.4) holds for $r=n+1$.

It follows similarly that $a_{i+r}=a_{i+r+\omega}$ for $r \leqq 0$. We thereby make use of the fact that two similar m-blocks must be preceded by symbols with equal values. One starts with the similarity relation (7.3) and proceeds inductively. We conclude that (a) has the period ω, and the proof is complete.

COROLLARY. *For a non-periodic trajectory $P(m+1)>P(m)$ for each $m>0$.*

There exist trajectories for which $P(n+1)=P(n)+1$, so that the relation $P(n+1)>P(n)$ of the corollary is in a sense the best possible. An example of a recurrent trajectory for which $P(n+1)=P(n)+1$ for each $n>0$ will be given in a later paper.

LEMMA 7.2. *If t is a trajectory which involves μ generating symbols and has no period less than ω, then*

$$P(n) \geqq n+\mu-1 \tag{7.5}$$

for all values n for which $P(n)<\omega$.

Relation (7.5) is clearly true when $n=1$. From the preceding theorem $P(m+1)>P(m)$ for each m for which $P(m)<\omega$. Relation (7.5) follows inductively.

THEOREM 7.4. *If t is a non-periodic trajectory involving μ generating symbols,*

$$P(n) \geqq n + \mu - 1.$$

LEMMA 7.3. *If t is a recurrent trajectory with no period less than ω, then*

$$R(n) \geqq P(n) + n \geqq 2n + \mu - 1 \tag{7.6}$$

for each n such that $P(n) < \omega$.

Suppose that $P(n) < \omega$ and set $R(n) = m$. Since t is recurrent, each m-block B contains each of the $P(n)$ n-blocks of t. Hence

$$m \geqq P(n) + n - 1. \tag{7.7}$$

We shall exclude the equality in (7.7).

Suppose B has the form $[r+1, r+m]$ in an indexed representation (a) of t. Suppose $m = P(n) + n - 1$. Then no two n-blocks of B are similar. In particular the initial n-block A of B does not appear twice in B. But $R(n) = m$ so that A appears in the I-block $[r+2, r+m+1]$, and since A appears but once in B, A appears with final index $r+m+1$. Since A appears with final index $r+n$ as well, $a_{r+n} = a_{r+m+1}$ for arbitrary r. Hence t has a period

$$m - n + 1 = [P(n) + n - 1] - n + 1 = P(n) < \omega.$$

This is contrary to the hypothesis that t has no period less than ω. Hence $m > P(n) + n - 1$, and (7.6) follows with the aid of (7.5).

Lemma 7.3 and Theorem 7.4 yield the following theorem.

THEOREM 7.5. *If t is a non-periodic recurrent trajectory involving μ generating symbols,*

$$R(n) \geqq P(n) + n \geqq 2n + \mu - 1, \tag{7.8}$$

where $R(n)$ and $P(n)$ are respectively the recurrency and permutation indices of t.

It will be convenient to set $P(0) = 0$ for each trajectory t.

THEOREM 7.6. *If a trajectory t involves μ generating symbols and has a minimum period ω,*

$$R(n) \leqq n + \omega - 1 \qquad (n = 1, 2, \cdots). \tag{7.9}$$

Moreover there exists a largest integer m such that $P(m) < \omega$. For such an m,

$$R(n) \geqq 2n + \mu - 1 \qquad (n = 1, 2, \cdots, m \leqq \omega - \mu), \tag{7.10}$$

$$R(n) = n + \omega - 1 \qquad (n = m+1, m+2, \cdots). \tag{7.11}$$

Relation (7.9) is immediate. To establish the existence of a largest integer m such that $P(m) < \omega$, observe that each integer $n > 0$ for which $P(n) < \omega$ satisfies the relations

$$2n + \mu - 1 \leqq R(n) \leqq n + \omega - 1,$$

by virtue of (7.6) and (7.9). It follows that m exists and is at most $\omega - \mu$. Relation (7.10) follows from (7.6).

To establish (7.11) let n be any integer $> m$. Then $P(n) \geqq \omega$ and

$$R(n) \geqq P(n) + n - 1 \geqq \omega + n - 1. \tag{7.12}$$

Relation (7.11) follows from (7.12) and (7.9).

Let (a) be an arbitrary I-trajectory. Let n be a positive integer, and let m be an integer, positive, negative or zero. For n and m fixed, let $H(m)$ be the minimum length of those blocks of (a) with initial index m which contain copies of all n-blocks of (a). If no such block with initial index m exists, let $H(m) = +\infty$. If, as m ranges over all integers, the set $H(m)$ is bounded, it is clear that the n-th recurrency index $R(n)$ of (a) exists and

$$R(n) = \max. H(m). \tag{7.13}$$

Let r be any integer such that $H(r+1)$ assumes the maximum of $H(m)$. The $R(n)$-block $[r+1, r+R(n)]$ will then be termed a *minimax n-covering* in (a) or in the trajectory t represented by (a). The length of a minimax n-covering B in t is $R(n)$ and the final n-block in B does not otherwise appear in B. These two properties are characteristic provided $R(n)$ exists.

We are led to the following theorem.

THEOREM 7.7. *If t is a trajectory for which $R(n+1)$ and $R(n)$ exist, $R(n+1) > R(n)$.*

To prove the theorem let B be a minimax n-covering in t. Let A be the final n-block in B and e the symbol following A. The $(n+1)$-block Ae does not appear in B since A has no second appearance in B. Hence $R(n+1) > R(n)$, as stated.

It may happen that $R(n+1) = R(n) + 1$ for infinitely many values of n as the recurrent trajectory of the following section shows.

8. The Morse recurrent trajectory. In this section we shall recall the definition of the non-periodic I-trajectory T first defined by Morse [2], p. 95, and shall give the explicit values of the corresponding recurrency function, denoting this function by $\rho(n)$.

Recall that the symbols α and β can be combined in any way to form admissible blocks. We here take α and β as 1 and 2 respectively. Set

$$(8.0)\qquad \begin{aligned} a_0 &= 1, & b_0 &= 2, \\ a_1 &= a_0 b_0, & b_1 &= b_0 a_0, \\ &\cdots & &\cdots \\ a_{n+1} &= a_n b_n, & b_{n+1} &= b_n a_n, \end{aligned}$$

understanding that the blocks $a_n b_n$ and $b_n a_n$ are to be expanded as blocks of 1's and 2's of length 2^{n+1}. The I-trajectory T shall have the representation

$$(8.1)\qquad \cdots c_{-2}c_{-1}c_0c_1c_2\cdots.$$

In T we set

$$(8.2)\qquad c_0c_1\cdots c_{2^n-1} \equiv a_n.$$

Note that (8.2) is consistent for all n, since a_m is an initial subblock of a_n for $m < n$. The symbols c_i are thereby defined for $i > 0$. To complete the definition of (8.1) we set

$$(8.3)\qquad c_{-r} = c_{r-1} \qquad (r > 0).$$

The subray

$$(8.4)\qquad R = c_0c_1c_2\cdots$$

of T thus starts as follows:

$$(8.5)\qquad 1221\ 2112\ 2112\ 1221\ 2112\ 1221\ 1221\ 2112\cdots.$$

Corresponding to an arbitrary block A, the block which consists of the symbols of A in inverse order will be denoted by A^{-1}. We note that

$$(8.6')\qquad a_n = a_n^{-1}, \quad b_n = b_n^{-1}, \qquad (n \text{ even})$$

$$(8.6'')\qquad a_n = b_n^{-1}, \quad b_n = a_n^{-1}, \qquad (n \text{ odd}).$$

Relations (8.6) are clearly true when $n = 0$, and their truth in general follows inductively from the relations

$$a^{-1}{}_{m+1} = (a_m b_m)^{-1} = b_m{}^{-1} a_m{}^{-1},$$
$$b^{-1}{}_{m+1} = (b_m a_m)^{-1} = a_m{}^{-1} b_m{}^{-1}.$$

From (8.6) we infer that the 2^n-block of (8.1) whose final index is -1 is $a_n{}^{-1} = a_n$ when n is even and $a_n{}^{-1} = b_n$ when n is odd. Thus the 2^{n+1}-block

whose initial index is -2^n is $a_n a_n$ when n is even and $b_n a_n$ when n is odd. Each block of T appears in such a 2^{n+1}-block for n sufficiently large. Since

$$a_{n+3} = a_n b_n b_n a_n b_n a_n a_n b_n,$$

each block of T appears in the subray R of (8.4). The recurrency indices of T (if they exist) are thus completely determined by this subray.

Let A be an arbitrary block of 1's and 2's. The block obtained from A by replacing 1 by 2 and 2 by 1 will be denoted by A'. An easy inductive proof shows that

$$a'_n = b_n, \qquad b'_n = a_n. \tag{8.7}$$

The trajectories T' and T_n. Let T' be a trajectory with an indexed representation (e) in which $e_i = c_i$ for $i \geqq 0$, and for $i < 0$ equals 1 or 2 according as c_i is 2 or 1 respectively. The 2^n-block of T whose final index is -1 is a_n or b_n according as n is even or odd. The corresponding 2^n-block of T' is then $a'_n = b_n$ or $b'_n = a_n$ according as n is even or odd. We infer that each block of T' has a copy in the subray R of (8.4). The recurrency indices of T' (if they exist) are then completely determined by R. Since this is also true of T we have proved the following lemma.

LEMMA 8.1. *The recurrency indices of T and T', if they exist, are equal and are uniquely determined by the ray R in* (8.4).

We introduce a trajectory T_n, $n \geqq 0$, whose generating symbols are a_n and b_n, and which is obtained from T or T' according as n is even or odd, by replacing 1 by a_n and 2 by b_n. It follows from the relations (8.0), (8.3) and (8.6) that an expansion of the symbols a_n and b_n in T_n into blocks of 1's and 2's will yield T. Unless otherwise stated T_n shall be regarded as a trajectory in the symbols a_n and b_n unexpanded as blocks of 1's and 2's. For example $a_n b_n$ will be regarded as a 2-block of T_n, or, if we choose, as a 2^{n+1}-block of T.

We shall make use of the following properties of the ray R of (8.4).

(a). *There are no more than two successive* 1's *or* 2's *in R.*

(b). *Each 5-block of R contains the blocks* 12 *and* 21.

(c). *Each 5-block of R contains either the block* 11 *or else the block* 22.

To establish (a) we note that each 3-block of R will be found in one of the expanded 2-blocks of T_1; that is, in one of the blocks

$$a_1 b_1 = 1221, \qquad a_1 a_1 = 1212,$$
$$b_1 a_1 = 2112, \qquad b_1 b_1 = 2121,$$

and no more than two successive 1's or 2's appear in these blocks.

To verify (b) observe that a 5-block B which did not contain 12 would contain 21 in its first 3-block by virtue of (a). Its final 3-block would then be 111 contrary to (a). We infer that B contains 12, and similarly 21.

To prove (c) we infer from (a) that there can be no more than two consecutive a_1's or b_1's in T_1. In particular R cannot contain the block 12121 since R would then contain the block $a_1a_1a_1$ or $b_1b_1b_1$ according as 12121 possessed an even or odd first index. Similarly R contains no block 21212, and (c) follows directly.

Statements (a), (b) and (c) are true if R is replaced by T or T' since each block in T or T' appears in the subray R.

We continue with the following lemma.

LEMMA 8.2. *The recurrency index $\rho(n)$ of the trajectory T has the values*

$$\rho(1) = 3, \qquad \rho(2) = 9, \qquad \rho(3) = 11.$$

That $\rho(1) = 3$ follows from (a). We continue by showing that $\rho(2) \leqq 9$. To that end consider a 5-block B of T_1. The block B will contain the blocks

$$a_1b_1 = 1221, \qquad b_1a_1 = 2112, \tag{8.8}$$

by virtue of (b) and either

$$a_1a_1 = 1212 \quad \text{or} \quad b_1b_1 = 2121, \tag{8.9}$$

by virtue of (c). The only 2-blocks in T are 12, 21, 11 and 22. If B_0 denotes the block B expanded in terms of 1's and 2's, we see that each 2-block of T appears in B_0 *interior* to B_0. Thus the end symbols of B_0 can be deleted and the resulting block B'_0 contains all 2-blocks of T. But any 9-block of T will contain some block of the nature of B'_0 so that $\rho(2) \leqq 9$.

That $\rho(2)$ is not less than 9 follows upon considering the 9-block of T with initial index 10. This block has the form

$$12\ 1221\ 211,$$

and will contain no block 11, if the final 1 is removed. Thus $\rho(2) = 9$.

To show that $\rho(3) = 11$, recall that there are no more than two successive 1's or 2's in T so that the only 3-blocks are

$$112, \quad 122, \quad 121, \quad 211, \quad 221, \quad 212. \tag{8.10}$$

Consider the block B_0 of the preceding proof. It contains the blocks (8.8) and at least one of the blocks (8.9). It follows that B_0 contains each 3-block (8.10). But any 11-block of T will contain a B_0, so that $\rho(3) \leqq 11$.

That $\rho(3) \geqq 11$ follows upon noting that the 11-block of T with initial index 3 has the form

$$1\,2112\,2112\,12,$$

and that removal of the final symbol from this block leaves no subblock of the form 212. Hence $\rho(3) = 11$, and the proof of the lemma is complete.

Let T'_1 be the trajectory obtained from T' by grouping the symbols of T' into 2-blocks with even first indices, regarding T'_1 as generated by the blocks a_1, b_1. It is readily seen that T'_1 is formed from the blocks a_1, b_1 exactly as T is formed from the symbols 1 and 2 respectively. We shall make use of this fact in proving the following lemma.

LEMMA 8.3. *No two r-blocks of $T(T')$ for which $r \geqq 4$ are equal unless their first indices are equal mod 2.*

It is sufficient to prove the lemma for the case $r = 4$. To that end observe that each 4-block of $T(T')$ as it stands is in some 6-block A of $T(T')$ with even first index. Let A_1 be the 3-block of $T_1(T'_1)$ determined by A. Since $\rho(3) = 11$, each 3-block of $T_1(T'_1)$ appears in the 11-block of $T_1(T'_1)$ with first index 0. The lemma follows upon verifying its truth for the 22-block of T with first index 0. For each 4-block of T appears in this 22-block with a first index which is unchanged mod 2.

The following lemma is useful.

LEMMA 8.4. *If $\rho(m+1)$ exists and $m \geqq 2$,*

$$\rho(2m) \leqq 2\rho(m+1) - 1.$$

We need only to show that each block B of T of length $2\rho(m+1) - 1$ contains a copy of each $2m$-block x of T. We distinguish two cases.

Case 1. *The first index of x is even.* Here x determines an m-block of T_1 and so appears in each expanded $\rho(m)$-block of T_1. Hence x appears in each $[2\rho(m) + 1]$-block of T. Thus $\rho(2m)$ exists and

$$2\rho(m) + 1 \geqq \rho(2m).$$

But

$$\rho(m+1) \geqq \rho(m) + 1,$$

in accordance with Theorem 7.7. Hence

$$2\rho(m+1) - 1 \geqq 2\rho(m) + 1 \geqq \rho(2m).$$

This completes the proof in Case 1.

Case 2. *The first index of x is odd.* Let c' and c'' be respectively the elements of T which precede and follow the above block B. If B has an odd first index, $c'B$ has an even first index and a length $2\rho(m+1)$ and so contains a copy of each $(m+1)$-block of T_1. Hence $c'B$ must contain a copy y of x. But the index of y is odd in accordance with Lemma 8.3, so that y does not include c'. Hence x appears in B. If B has an even first index, Bc'' contains a copy y of x. But the index of c'' is odd, while the final index of x is even. Hence y does not include c'' and lies in B.

The proof of the lemma is complete.

Since $\rho(3)$ exists it follows from the lemma that $\rho(4)$ exists, then $\rho(6)$, $\rho(10)$, etc. But if $\rho(n)$ exists, $\rho(m)$ exists for $m < n$. We conclude that $\rho(n)$ exists for all n.

We come to a principal theorem.

THEOREM 8.1. *For $m \geqq 2$ the recurrency function $\rho(m)$ of T satisfies the relations*

$$\rho(2m) = 2\rho(m+1) - 1, \tag{8.11}$$

$$\rho(2m+1) = 2\rho(m+1). \tag{8.12}$$

Proof of (8.11). It will be sufficient to show that the equality prevails in Lemma 8.4. To that end let g be a minimax $(m+1)$-covering in T_1 as defined in § 7. The length of g is $\rho(m+1)$, and the final $(m+1)$-block of g appears only once in g. Let x be an arbitrary $2m$-block of T of odd first index, and let y be an $(m+1)$-block of T_1 which when expanded contains x. When $m \geqq 2$, x appears only as an interior block of the expanded y in accordance with Lemma 8.3. Since a 2-block of T of even index is uniquely determined by either of its symbols we see that y is uniquely determined by x. It thus requires copies of all $(m+1)$-blocks of T_1 to cover all $2m$-blocks of T of odd index. In particular it will require a copy of the final $(m+1)$-block of g. As a block of T, g can then be shortened on the right by at most one symbol and, when expanded, still cover all $2m$-blocks of T. The equality accordingly holds in Lemma 8.4, and the proof of (8.11) is complete.

Proof of (8.12). Recall that $\rho(2m+1)$ exceeds $\rho(2m)$. It follows then from (8.11) that

$$\rho(2m+1) \geqq 2\rho(m+1). \tag{8.13}$$

It remains to show that the equality prevails in (8.13). To that end let B be a $2\rho(m+1)$-block of T, and let c' and c'' be respectively the symbols pre-

ceding and following B. Let x be an arbitrary $(2m+1)$-block of T. We distinguish between several cases:

Case 1. The index of c' is even, that of x odd.
Case 2. The index of c' is even, that of x even.
Case 3. The index of c' is odd.

Case 1. The block x appears in some expanded $(m+1)$-block of T_1. Hence x has a copy y in $c'B$, for the first $2\rho(m+1)$-block of $c'B$ determines a $\rho(m+1)$-block of T_1. But the index of y is odd so that y does not include c' and thus lies in B.

Case 2. Here c'' has an odd index and the last $2\rho(m+1)$-block of Bc'' determines a $\rho(m+1)$-block of T_1. Hence x has a copy y in Bc''. But the final index of y is even so that y lies in B.

Case 3. In this case B determines a $\rho(m+1)$-block of T_1, hence contains a copy of each $(m+1)$-block of T_1, and accordingly a copy of x.

Thus a copy of x appears in B in all cases. The equality accordingly prevails in (8.13), and the proof of (8.12) is complete.

With the aid of the preceding theorem we can derive a general formula for $\rho(n)$ when $n \geqq 3$. To that end we first show that integers $n \geqq 3$ can be represented in the form

$$n = 2^r + p, \qquad (p = 2, 3, \cdots, 2^r + 1), \tag{8.14}$$

where $r \geqq 0$ and r and p are uniquely determined by n. The smallest and largest values of n given by (8.14) for a fixed r are respectively

$$2^r + 2, \qquad 2^{r+1} + 1,$$

and when r is replaced by $r+1$ the smallest value of n is

$$2^{r+1} + 2.$$

Thus as r ranges over the values $0, 1, 2, \cdots$ and p over the corresponding values listed in (8.14), n takes on each of the integers $3, 4, \cdots$ once and only once. We continue with the proof of the following theorem.

THEOREM 8.2. *For values of n of the form* (8.14) *the recurrency function $\rho(n)$ of T has the value*

$$\rho(n) = 10 \cdot 2^r + p - 1. \tag{8.15}$$

When $r = 0$ in (8.14) $p = 2$ and $n = 3$. We have seen that $\rho(3) = 11$,

so that (8.15) holds when $n = 3$. When $r = 1$, we see that $p = 2$ and 3 while $n = 4$ and 5 respectively in (8.14). Relation (8.15) affirms that $\rho(4) = 21$ and $\rho(5) = 22$, and these values are correct as follows from (8.11) and (8.12) respectively upon setting $m = 2$. We proceed inductively assuming that (8.15) holds for a fixed r and the related values of p, and seek to prove that

$$\rho(\mu) = 10 \cdot 2^{r+1} + q - 1 \tag{8.16}$$

where

$$\mu = 2^{r+1} + q \qquad (q = 2, \cdots, 2^{r+1} + 1). \tag{8.17}$$

We distinguish between the cases where q is even and q is odd.

Case 1. *q even.* Here q has the form $2s$ and

$$\mu = 2(2^r + s).$$

Upon setting $m = 2^r + s$ in (8.11) we find that

$$\rho(\mu) = 2\rho(2^r + s + 1) - 1. \tag{8.18}$$

Upon referring to (8.17) we see that $s + 1$ ranges over the values

$$2, 3, \cdots, 2^r + 1. \tag{8.19}$$

When $n = 2^r + s + 1$, $p = s + 1$ in (8.14) and the range (8.19) is also the range prescribed in (8.14). We can accordingly use our inductive hypothesis to infer that

$$\rho(2^r + s + 1) = 10 \cdot 2^r + p - 1 = 10 \cdot 2^r + s. \tag{8.20}$$

Upon returning to (8.18) we conclude that

$$\rho(\mu) = 2(10 \cdot 2^r + s) - 1 = 10 \cdot 2^{r+1} + q - 1.$$

Relation (8.16) thus holds when q is even.

Case 2. *q odd.* Let q and μ be values of q and μ given by (8.17) with q odd. Corresponding to q and μ there exists a pair of even integers q_1, μ_1 again satisfying (8.17) with $\mu_1 = 2m$ and

$$q = q_1 + 1, \qquad \mu = 2m + 1.$$

Upon using the result established under Case 1,

$$\rho(\mu_1) = \rho(2m) = 10 \cdot 2^{r+1} + q_1 - 1.$$

But

$$\rho(2m + 1) = \rho(2m) + 1$$

by virtue of (8.11) and (8.12), so that in Case 2

$$\rho(\mu) = 10 \cdot 2^{r+1} + q_1 = 10 \cdot 2^{r+1} + q - 1.$$

Thus (8.16) holds in Case 2, and the proof of the theorem is complete.

This explicit determination of $\rho(n)$ is the first such determination for a non-periodic, recurrent trajectory.

The manner in which $\rho(n)$ varies with n is illustrated by the following table:

$n = 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18,$
$\rho(n) = 3, 9, 11, \underline{21, 22,} \underline{41, 42, 43, 44,} \underline{81, 82, 83, 84, 85, 86, 87, 88,} 161,$

where the groups underlined are related in an obvious manner. For a given r in (8.14) the smallest and largest values of $\rho(n)$ are respectively

$$10 \cdot 2^r + 1, \qquad 11 \cdot 2^r.$$

We see that for a fixed r,

$$10 > \frac{10 \cdot 2^r + 1}{2^r + 2} \geqq \frac{\rho(n)}{n} = \frac{10 \cdot 2^r + p - 1}{2^r + p} \geqq \frac{11 \cdot 2^r}{2^{r+1} + 1}.$$

We are thus led to the following theorem.

THEOREM 8.3. *If $\rho(n)$ is the recurrency function of T,*

$$\text{(8.21)} \qquad \liminf_{n\to\infty} \frac{\rho(n)}{n} = \frac{11}{2}, \qquad \limsup_{n\to\infty} \frac{\rho(n)}{n} = 10,$$

and 10 and 11/2 are the limits respectively of $\rho(n)/n$ for sequences of values of n given by

$$n = 2^r + 2, \qquad n = 2^{r+1} + 1, \qquad (r = 0, 1, \cdots).$$

If e is a prescribed positive constant and r is a sufficiently large positive integer, the ratios

$$\frac{\rho(n)}{n} = \frac{10 \cdot 2^r + p - 1}{2^r + p} \qquad (p = 2, 3, \cdots, 2^r + 1)$$

assume values within e of each value k between the limits (8.21). There accordingly exists a subsequence of values of n for which

$$\lim_{n\to\infty} \frac{\rho(n)}{n} = k.$$

For all values of n, $\rho(n) < 10n$, while for each positive e and sufficiently large values of n, $2\rho(n) > (11 - e)n$.

That the trajectory T is not periodic follows from Theorem 7.6.

In § 4 a trajectory has been defined as almost periodic if it admits a

relatively dense set of ϵ-translations for any positive ϵ. We show that a recurrent trajectory is not necessarily almost periodic by proving that the Morse recurrent trajectory admits only the trivial ϵ-translation if $\epsilon < 1/4$. To that end we first prove two lemmas.

LEMMA 8.5. *If $\cdots \delta_1\delta_0\delta_1 \cdots$ is a sequence of the numbers 0 or 1, and p and κ are integers, $\kappa > 0$, then*

$$\limsup_{n\to\infty} \frac{1}{2n+1} \sum_{i=-n}^{n} \delta_i = \limsup_{n\to\infty} \frac{1}{2n+1} \sum_{i=-n}^{n+p} \delta_i = \limsup_{m\to\infty} \frac{1}{2\kappa m+1} \sum_{i=-\kappa m}^{\kappa m} \delta_i.$$

The first equality follows at once from the relationship

$$\lim_{n\to\infty} \left| \frac{1}{2n+1} \sum_{i=-n}^{n} \delta_i - \frac{1}{2n+1} \sum_{i=-n}^{n+p} \delta_i \right| \leqq \lim_{n\to\infty} \frac{p}{2n+1} = 0.$$

As to the second, let $n = \kappa m + r$, $0 \leqq r < \kappa$. Then

$$\frac{1}{2n+1} \sum_{i=-n}^{n} \delta_i = \frac{1}{2n+1} \sum_{i=-\kappa m}^{\kappa m} \delta_i + \frac{1}{2n+1} \sum_{i=-\kappa m-r}^{-\kappa m-1} \delta_i + \frac{1}{2n+1} \sum_{i=\kappa m+1}^{\kappa m+r} \delta_i,$$

where the last two sums are not present if $r = 0$. Each of these last two sums approaches 0 as n becomes infinite. It follows that

$$\limsup_{n\to\infty} \frac{1}{2n+1} \sum_{i=-n}^{n} \delta_i = \limsup_{n\to\infty} \frac{1}{2n+1} \sum_{i=-\kappa m}^{\kappa m} \delta_i$$
$$= \limsup_{n\to\infty} \frac{2\kappa m+1}{2n+1} \cdot \frac{1}{2\kappa m+1} \sum_{i=-\kappa m}^{\kappa m} \delta_i = \limsup_{m\to\infty} \frac{1}{2\kappa m+1} \sum_{i=-\kappa m}^{\kappa m} \delta_i.$$

The proof of the lemma is complete.

We recall that T_n is the trajectory with generating symbols a_n and b_n obtained from T or T', according as n is even or odd, by replacing 1 by a_n and 2 by b_n and that an expansion of the symbols a_n and b_n in T_n into blocks of 1's and 2's will yield T.

LEMMA 8.6. *If $r = 2^p q$, $p \geqq 0$, and (c) and (d) are I-trajectories representing T and T_p, respectively,*

$$[D_r(c), (c)] = [D_q(d), (d)].$$

It follows from Lemma 4.1 that the value of $[D_r(c), (c)]$ is independent of the particular indexed representation (c) of T which we use. We shall find it convenient to take (c) as the I-trajectory (8.1). Then $D_r(c)$ is an I-trajectory (c') in which $c'_i = c_{i-r}$. Similarly we can suppose that in (d), d_i in expanded form is given by

$$d_i = c_{i2^p} \cdots c_{(i+1)2^p-1},$$

for (d) is then an indexed representation of T_p. Since $r = 2^p q$, $D_q(d)$ has the form (d') where $d'_i = d_{i-q} = c_{i2^p-r} \cdots c_{(i+1)2^p-r-1}$, in expanded form.

Since d_i and d'_i are either a_p or b_p, it follows from (8.7) that in their expanded form, d_i and d'_i are either identical, or one is obtained from the other by replacing 1 by 2 and 2 by 1. If we set

$$\{d'_i - d_i\} = \sum_{j=i2^p}^{(i+1)2^p-1} | c'_j - c_j |,$$

it follows that

$$\{d'_i - d_i\} = 2^p | d'_i - d_i |.$$

With the aid of Lemma 8.5, we then see that

$$\begin{aligned}
[D_r(c), (c)] &= \limsup_{n\to\infty} \frac{1}{2n+1} \sum_{i=-n}^{n} | c'_i - c_i | \\
&= \limsup_{m\to\infty} \frac{1}{2\cdot 2^p m + 1} \sum_{i=-m2^p}^{m2^p} | c'_i - c_i | \\
&= \limsup_{m\to\infty} \frac{1}{2^{p+1}m + 1} \sum_{i=-m2^p}^{m2^p+2^p-1} | c'_i - c_i | \\
&= \limsup_{m\to\infty} \frac{1}{2^{p+1}m + 1} \sum_{i=-m}^{m} \{d'_i - d_i\} \\
&= \limsup_{m\to\infty} \frac{2^p}{2^{p+1}m + 1} \sum_{i=-m}^{m} | d'_i - d_i | \\
&= \limsup_{m\to\infty} \frac{1}{2m+1} \sum_{i=-m}^{m} | d'_i - d_i | = [D_r(d), (d)].
\end{aligned}$$

The proof of the lemma is complete.

We come to the proof of the desired theorem.

THEOREM 8.4. *The Morse recurrent trajectory admits only the trivial ϵ-translation 0 if $\epsilon < 1/4$.*

It is sufficient to show that

$$[D_r(c), (c)] \geqq \tfrac{1}{4}, \qquad r \neq 0, \tag{8.22}$$

where (c) is again the indexed representation (8.1) of the Morse recurrent trajectory.

Case 1. *r odd.* Let

$$\sigma_n = \sum_{i=n}^{n+3} | c'_i - c_i | = \sum_{i=n}^{n+3} | c_{i-r} - c_i |.$$

Since r is odd, the first indices of $c_{n-r}c_{n-r+1}c_{n-r+2}c_{n-r+3}$ and $c_nc_{n+1}c_{n+2}c_{n+3}$ are not

congruent mod 2. It follows from Lemma 8.3 that these two 4-blocks cannot be similar, so that $\sigma_n \geqq 1$. Applying Lemma 8.5, we find that

$$\begin{aligned}[D_r(c), (c)] &= \limsup_{n\to\infty} \frac{1}{2n+1} \sum_{i=-n}^{n} | c'_i - c_i | \\ &= \limsup_{n\to\infty} \frac{1}{8n+1} \sum_{i=-4n}^{4n+3} | c'_i - c_i | \\ &= \limsup_{n\to\infty} \frac{1}{8n+1} \sum_{i=-n}^{n} \sigma_{4i}.\end{aligned}$$

Since $\sigma_n \geqq 1$, it follows that $\sum_{i=-n}^{n} \sigma_{4i} \geqq 2n+1$, and hence

$$(8.23) \quad [D_r(c), (c)] = \limsup_{n\to\infty} \frac{1}{8n+1} \sum_{i=-n}^{n} \sigma_{4i} \geqq \lim_{n\to\infty} \frac{2n+1}{8n+1} = \tfrac{1}{4}.$$

We have proved (8.22) when r is odd.

Let $(\bar{c})$ be the I-trajectory obtained from (c) by replacing c_n, $n < 0$, by 1 if c_n is 2, and by 2 if c_n is 1. Then $(\bar{c})$ is an indexed representation of the trajectory T'. Since Lemma 8.3 holds for T', as well as T, a proof similar to that given for (c) yields the relation

$$(8.24) \qquad [D_r(\bar{c}), (\bar{c})] \geqq \tfrac{1}{4}, \qquad r \text{ odd}.$$

Case 2. r even. Since $r \neq 0$, $r = 2^p q$, where q is odd. From Lemma 8.6,

$$(8.25) \qquad [D_r(c), (c)] = [D_q(d), (d)].$$

The symbols of (d) are a_p and b_p, and if in (d) we set $a_p = 1$, $b_p = 2$, we obtain (c) or $(\bar{c})$ according as p is even or odd. Since q is odd, (8.23) or (8.24) applied to the right side of (8.25) yields

$$[D_q(d), (d)] \geqq \tfrac{1}{4}.$$

Combining this inequality with (8.25) we conclude that

$$[D_r(c), (c)] \geqq \tfrac{1}{4}, \qquad r \text{ even}.$$

The proof of the theorem is complete.

The following corollary is an immediate consequence of the preceding theorem.

COROLLARY. *The Morse recurrent trajectory is not almost periodic.*

Upon distinguishing recurrent trajectories according to whether they are almost periodic or not, we can state that recurrent trajectories of the type of

Morse are the only ones so far defined which are not almost periodic. In particular, the recurrent trajectories defined by Birkhoff ([4], p. 22) are almost periodic. The proof of this and the further development of the theory of almost periodic trajectories will be given in a later paper by the authors.

9. The derivation of recurrent trajectories. Having exhibited a non-periodic recurrent trajectory T, we shall describe four methods of deriving a recurrent trajectory from T, termed respectively derivation by projection, reduction, association and substitution. We also introduce a general method of derivation which includes each of the preceding. If the given trajectory consists exclusively of the free symbols α and β, these derived trajectories will all be admissible. In general they may or may not be admissible. If the given trajectory T is recurrent, the derived trajectories will be recurrent although usually differing in high degree from T. We suppose that T has an indexed representation (c) in terms of generating symbols $a_1, \cdots, a_\mu$.

Projection. The trajectory obtained from (c) by omitting a_k whenever it occurs in (c) will be said to be the k-th projection of (c).

Reduction. If the generating symbols consist of a finite set of integers, the trajectory obtained by reducing the elements $c_j \bmod p$ will be said to be derived by reduction mod p.

Association. Let m be a positive integer at least 2, and let B_i be the m-block of (c) whose first index is i. There is at most a finite set of different blocks B_i. We introduce the I-trajectory

$$\cdots B_{-1}B_0B_1\cdots, \tag{B}$$

regarding the blocks B_i as the generating symbols.

Substitution. Let $A_1, \cdots, A_\mu$ be s-blocks of the generating symbols, $s > 0$. The trajectory (c^*) of a_i's obtained by replacing a_i in (c) by A_i and expanding will be said to be derived from (c) by block substitution.

Generalized substitution. Let $r_1, \cdots, r_m$ be a fixed set of integers, positive, negative, or zero. Let $\phi(x_1, \cdots, x_m)$ be a single-valued function (either numerical or symbolic) of its arguments x_i as each x_i ranges over the set $(a_1, \cdots, a_\mu)$ of generators of (c). We include the convention that a null set may be one of the symbolic values of ϕ. We introduce the symbols

$$E_i = \phi[c_{i+r_1}, c_{i+r_2}, \cdots, c_{i+r_m}],$$

and thereby define an I-trajectory (E). The number of different symbols E_i is finite. We add the convention that a null set E_i is to be dropped from the trajectory (E). Derivation by generalized substitution includes all of the preceding types of derivation. We shall illustrate this fact for the case of derivation by association. Derivation by association is obtained if we set

$$\phi(x_1, \cdots, x_m) = (x_1, \cdots, x_m),$$

and understand that

$$(r_1, \cdots, r_m) = (0, 1, \cdots, m-1).$$

The symbol E_i reduces to the block

$$B_i = c_i c_{i+1} \cdots c_{i+m-1},$$

as in the case of derivation by association.

We shall illustrate these processes further by deriving two recurrent trajectories from the trajectory T of the preceding section. We shall use the method of association letting $m = 2$. The generating symbols are the 2-blocks of T, namely 11, 12, 22, 21. We shall denote these blocks respectively by 1, 2, 3, 4 and write a few terms of the resulting trajectory. The symbols with indices $i \geqq 0$ written above the corresponding symbols of T begin as follows:

2342412341242342412

12121122112122121122.

Upon reducing the first of these trajectories mod 2 we obtain a trajectory which contains

0100010101000100010.

This trajectory is distinct from T since it has sequences of three equal symbols. These new trajectories are recurrent as we shall see.

A trajectory T' derived from a recurrent trajectory T by projection or reduction is clearly recurrent. The trajectory T' may however be periodic when T is non-periodic. The facts for the case of derivation by association are otherwise, as the following theorem shows.

THEOREM 9.1. *If (c') is a recurrent I-trajectory derived from a recurrent I-trajectory (c) by association in r-blocks, and $R'(n)$ and $R(n)$ are the recurrency functions of (c') and (c) respectively, then*

$$R'(n) = R(n+r-1) - r + 1. \tag{9.1}$$

Moreover (c') *is periodic if and only if* (c) *is periodic.*

A block $[k+1, k+n]$ of (c') is determined by the $(n+r-1)$-block $[k+1, k+n+r-1]$ of (c) and conversely. We term these blocks *corresponding*. Each $(n+r-1)$-block of (c) appears in each $R(n+r-1)$-block H of (c). But H corresponds in the above sense to a block (c') of length

$$R(n+r-1) - (r-1), \tag{9.2}$$

while an $(n+r-1)$-block of (c) corresponds to an n-block of (c'). Hence $R'(n)$ is at most the integer (9.2).

To show that $R'(n)$ is at least the integer (9.1), let H be a minimax $(n+r-1)$-covering in (c), that is an $R(n+r-1)$-block of (c) whose final $(n+r-1)$-block appears just once in H. Let K be the block of (c') corresponding to H. The final n-block of K corresponds to the final $(n+r-1)$-block of H and appears just once in K, so that $R'(n)$ is at least the integer (9.1), and the proof of (9.1) is complete.

If (c) is non-periodic, $R(n) > 2n$ in accordance with Theorem 7.5. It follows then from (9.1) that

$$R'(n) > 2n + r - 1.$$

The I-trajectory (c') is accordingly non-periodic, and the proof of the theorem is complete.

The preceding theorem leads to the following extension of Theorem 7.5.

THEOREM 9.2. *If* $P(r)$ *is the number of different* r*-blocks in a recurrent non-periodic* I*-trajectory* (c)*, then for* $m \geqq r$*,*

$$R(m) \geqq 2m + P(r) - r. \tag{9.3}$$

Let (c') be derived from (c) by association in r-blocks. If $R'(n)$ is the recurrency function of (c'), it follows from Theorem 7.5 that

$$R'(n) \geqq 2n + P(r) - 1.$$

But from the preceding theorem we then see that

$$R(n+r-1) - r + 1 = R'(n) \geqq 2n + P(r) - 1,$$

from which (9.3) follows upon setting $m = n + r - 1$.

Before coming to the next theorem we shall need the following definition.

A definition. Let there be given a finite set S of q-blocks $B_1, \cdots, B_p$ of a set of generating symbols. Let n be a positive integer less than q. A least integer $r(n) \leqq q$ (if such exists) such that each n-block appearing in a block of the set S appears in each $r(n)$-subblock of each of the blocks B_j will be called the n-th *recurrency index* of the set S. For example, let the blocks B_i be the blocks

$$(9.4) \qquad B_1 = 1221, \quad B_2 = 2112, \quad B_3 = 1212, \quad B_4 = 2121.$$

These blocks have a recurrency index $r(1) = 3$. The recurrency index $r(2)$ does not exist.

The following theorem concerns the I-trajectory (c^*) derived from the I-trajectory (c) by substitution. The I-trajectory (c^*) is formed as described above by substituting the respective s-blocks A_i for the generating symbols a_i appearing in (c). These blocks A_i are then expanded to form the I-trajectory (c^*) of the symbols a_i. We shall apply the preceding definition of the recurrency index of a set of blocks to the set S of $2s$-blocks A_iA_j. Each m-block of (c^*) for which $m \leqq s+1$ lies in one of these $2s$-blocks of (c^*) so that in the case where the set S has a recurrency index $r(m) \leqq s+1$ it is clear that the recurrency index $R^*(m)$ of (c^*) exists and is at most $r(m)$. This is affirmed in part (a) of the theorem. Part (b) conditions $R^*(n)$ on the lower side instead of on the upper side. This theorem is useful in the following section.

THEOREM 9.3. *Let (c^*) be an I-trajectory derived from an I-trajectory (c) by substitution as described above.*

(a) *If the set of $2s$-blocks A_iA_j of (c^*) has an m-th recurrency index $r(m) \leqq s+1$, then*

$$(9.5) \qquad R^*(m) \leqq r(m).$$

(b) *If some block A_i appears r times consecutively in (c^*) with $r > 1$, and if (c^*) does not have the period s, then*

$$(9.6) \qquad R^*(s+1) > sr,$$

or else $R^*(s+1)$ fails to exist.

Statement (a) has already been proved. We therefore turn to the proof of (b).

It is assumed that the block $B = A_i \cdots A_i$, of length sr, appears in (c^*). Since the length of A_i is s, any $(s+1)$-block of B has equal terminal elements.

If $R^*(s+1)$ exists and $R^*(s+1) \leqq sr$, every $(s+1)$-block of (c^*) has a copy in B. Thus every $(s+1)$-block of (c^*) has equal terminal elements and (c^*) has the period s, contrary to hypothesis.

The proof of the theorem is complete.

The following theorem has a special interest in view of the mode of formation of the Morse recurrent trajectory. The trajectory H thereby affirmed to exist is not necessarily admissible. It will be admissible if the only symbols employed are α and β.

THEOREM 9.4. *Corresponding to an arbitrary recurrent ray*

$$R = a_0 a_1 a_2 \cdots,$$

there exists a recurrent trajectory H with R as subray and with the recurrency indices of R. The trajectory H is in general not uniquely determined by R.

Let B_r be the initial r-block of R, $r = 1, 2, \cdots$. The block B_r appears later in R immediately preceded by some r-block A_r. Of the blocks $A_1, A_2, \cdots$ there exists an infinite subset S_1 with last symbol in common. We take a_{-1} as this last symbol. Of the blocks of S_1 there exists an infinite subset S_2 with the second from the last symbol in common. We take a_{-2} as this common symbol. Proceeding in this way we define sets $S_3, S_4, \cdots$ and symbols $a_{-3}, a_{-4}, \cdots$ respectively. That the resulting trajectory H is recurrent and has the same recurrency indices as R follows from the fact that each block of H appears in R.

That H is not uniquely determined is shown by the trajectories T and T' of § 8 which have the recurrent ray R of (8.4) as a common subray.

10. Deferred recurrence. In this section we shall exhibit a recurrent trajectory E for which $R(n)$ exceeds a prescribed function $f(n)$ for infinitely many values of n but which is such that the limit inferior of $R(n)/n$ is finite. That a recurrent trajectory exists for which $R(n)$ exceeds a prescribed function $f(n)$ for all values of n has already been shown by Robbins ([1], Theorem 2). Our example shows the possibility of extreme variability in the indices $R(n)$. The trajectories E here defined form an extended class of which the trajectory T of § 8 is a member.

The generating symbols shall be the free symbols α and β. Let $r_0, r_1, \cdots$ be an arbitrary set of positive integers. We shall make use of powers of blocks, understanding that these powers are to be formally expanded. Set

$$\begin{aligned} A_0 &= \alpha, & B_0 &= \beta, \\ A_1 &= A_0^{r_0} B_0^{r_0}, & B_1 &= B_0^{r_0} A_0^{r_0}, \\ &\cdots\cdots & &\cdots\cdots \\ A_{n+1} &= A_n^{r_n} B_n^{r_n}, & B_{n+1} &= B_n^{r_n} A_n^{r_n}. \end{aligned}$$

Let s_n be the length of A_n and B_n. Observe that

$$s_{n+1} = 2s_n r_n. \tag{10.1}$$

Let (c) be an I-trajectory such that the initial s_n-block of the ray $c_0c_1c_2\cdots$ is A_n for each n. This definition is self consistent since A_{n-1} is an initial block of A_n. We further set

$$c_{-m-1} = c_m \qquad (m = 0, 1, \cdots). \tag{10.2}$$

Proceeding inductively we see that

$$\begin{aligned} A_n^{-1} &= A_n, \quad & B_n^{-1} &= B_n, \quad & (n \text{ even}) \\ A_n^{-1} &= B_n, \quad & B_n^{-1} &= A_n, \quad & (n \text{ odd}). \end{aligned}$$

Hence the s_n-block with final index -1 is A_n when n is even and B_n when n is odd. If the symbols of (c) are grouped into s_r-blocks with first indices $\equiv 0 \bmod s_r$, one obtains an I-trajectory E_r with generating symbols A_r and B_r.

If the integers r_i are all 1 and $\alpha = 1$, $\beta = 2$, (c) yields the recurrent trajectory T of § 8. We shall make use of the recurrency function $\rho(n)$ of T. The trajectory defined by (c) will be denoted by E. It is uniquely determined by the symbols α and β and the integers r_i.

THEOREM 10.1. *The trajectory E is recurrent and non-periodic with a recurrency function $R^*(n)$ such that*

$$s_{n+1} \leqq R^*(s_n + 1) \leqq 2s_{n+3}. \tag{10.3}$$

Let B be an arbitrary $(s_n + 1)$-block of E. The block B appears in some expanded 2-block of E_n. But each such 2-block of E_n appears in both A_{n+3} and B_{n+3} as is easily seen. The latter two blocks are s_{n+3}-blocks of E and one of them at least is in each $2s_{n+3}$-block of E. Hence the right-hand condition in (10.3) is valid. It follows that E is recurrent.

The left condition in (10.3) follows from (b) in Theorem 9.3. For the block A_n appears $2r_n$ consecutive times in E. Moreover E does not have the period s_n since E contains the block A_nB_n and B_n is different from A_n. From (9.6) we then find that

$$R^*(s_n + 1) \geqq 2r_n s_n = s_{n+1},$$

and the proof of (10.3) is complete. Moreover we see that

$$R^*(s_n + 1) \geqq 2s_n$$

and it follows from Theorem 7.6 that E cannot be periodic.

COROLLARY. *If $f(n)$ is an arbitrary positive integral function, it is possible to choose the r_n's or any subset of the r_n's so that for corresponding values of n,*

$$R^*(s_n + 1) > f(s_n + 1). \tag{10.4}$$

Recall that $s_{n+1} = 2s_n r_n$. In order that (10.4) may hold we have merely to take r_n so large that

$$s_{n+1} > f(s_n + 1)$$

and (10.4) will follow from (10.3).

We continue with the following theorem.

THEOREM 10.2. *If the sequence of integers r_i used in defining E contains arbitrarily long blocks of 1's,*

$$\liminf_{n\to\infty} \frac{R^*(n)}{n} \leqq \frac{11}{2}. \tag{10.5}$$

If $\rho(m)$ is the recurrency function of the trajectory T of § 8, there exists a sequence of integers m_i becoming infinite with i and such that

$$\liminf_{i\to\infty} \frac{\rho(m_i)}{m_i} = \frac{11}{2}. \tag{10.6}$$

We suppose integers m_i so chosen. The sequence of integers r_i contains arbitrarily long blocks of 1's. It follows that there exists a sequence of positive integers

$$h_1 < k_1 < h_2 < k_2 < \cdots, \tag{10.7}$$

such that for each i,

$$r_{h_i} = r_{h_i+1} = \cdots = r_{k_i} = 1. \tag{10.8}$$

Moreover these integers can be chosen so that

$$\mu_i = k_i - h_i > \rho(m_i).$$

We shall prove the following statement.

(α). *The m_i-th recurrency index of E_{h_i} is at most the recurrency index $\rho(m_i)$.*

The trajectory E_{h_i} is generated from the blocks A_{h_i}, B_{h_i}. These blocks combine to form the blocks A_{k_i}, B_{k_i} in exactly the way in which the symbols 1 and 2 combine to form the blocks a_{μ_i} and b_{μ_i} of the Morse trajectory T. In E_{h_i} the blocks

$$A_{k_i}B_{k_i}, \quad A_{k_i}A_{k_i}, \quad B_{k_i}A_{k_i}, \quad B_{k_i}B_{k_i}, \tag{10.9}$$

regarded as blocks of the symbols A_{h_i} and B_{h_i}, accordingly have an m_i-th recurrency index $\rho(m_i)$. For

$$\rho(m_i) < k_i - h_i < 2^{k_i - h_i} = 2^{\mu_i},$$

and 2^{μ_i} is the length of the blocks A_{k_i} and B_{k_i} regarded as blocks of the symbols A_{h_i}, B_{h_i}. It follows from (a) of Theorem 9. 3 that E_{h_i} has an m_i-th recurrency index at most $\rho(m_i)$, and the proof of (α) is complete.

We now regard E as derived from E_{h_i} by expansion of the blocks A_{h_i} and B_{h_i} of E_{h_i} into blocks of α's and β's. Set

$$n_i = s_{h_i}(m_i - 1). \tag{10.10}$$

An arbitrary n_i-block of E is found in some expanded m_i-block of E_{h_i}, and hence in each expanded $\rho(m_i)$-block of E_{h_i}, and finally in each $[s_{h_i}\rho(m_i) + s_{h_i}]$-block of E. Hence

$$R(n_i) \leqq s_{h_i}[\rho(m_i) + 1]. \tag{10.11}$$

It follows from (10. 10), (10. 11) and (10. 6) that

$$\liminf_{i\to\infty} \frac{R(n_i)}{n_i} \leqq \liminf_{i\to\infty} \frac{\rho(m_i) + 1}{m_i - 1} = \frac{11}{2},$$

and the proof of the theorem is complete.

The preceding theorems and corollary combine to give use the following.

THEOREM 10. 3. *If $f(n)$ is an arbitrary positive integral function of n, then for suitable choices of the integers r_i defining E, the recurrency function $R^*(n)$ of E exceeds $f(n)$ for infinitely many values of n while*

$$\liminf_{n\to\infty} \frac{R^*(n)}{n} \leqq \frac{11}{2}.$$

11. Transitive rays and their ergodic functions. We shall concern ourselves with rays of the form

$$c_0c_1c_2 \cdots.$$

Such a ray will be termed transitive if it contains a copy of each admissible subblock. Transitive rays exist. If $A_1, A_2, \cdots$ is an enumeration of admissible blocks, the ray

$$X = A_1e_1A_2e_2 \cdots$$

is admissible provided the symbols e_i are successively chosen in a suitable way. The ray X is clearly transitive. Let R be an arbitrary ray and let n be a positive integer. Suppose an initial r-block B of R contains each admissible n-block while no initial subblock of B has this property. The integer r will then be termed the n-th *transitivity index* $\phi(n)$ of R. If R is transitive, the n-th transitivity index of R exists for each value of n and will be termed the *ergodic function* $\phi(n)$ of R.

A block $H(n)$ of minimum length containing each admissible n-block will be termed a *minimum n-covering*. The length $\theta(n)$ of $H(n)$ will be termed the n-th *covering index*. In the case where the generating symbols are all free, Martin [1] has shown that for each n there exists a minimum n-covering in which no n-block is repeated. We shall presently extend Martin's result to the case where successive inverses are prohibited. We are concerned here with the existence of transitive rays whose ergodic function is in some sense relatively small. Noting that each ergodic function $\phi(n)$ satisfies the condition

$$\phi(n) \geqq \theta(n),$$

we are led to the question, does there exist a transitive ray such that $\phi(n) = \theta(n)$ for each n? The answer is no, in general, as we shall see.

We begin with the case where there are just two generating symbols, namely the free symbols α and β. In this case it follows from Martin's theorem that

$$\theta(n) = 2^n + n - 1.$$

The only minimum 1-coverings are $\alpha\beta$ and $\beta\alpha$. The only minimum 2-covering which starts with $\alpha\beta$ is $\alpha\beta\beta\alpha\alpha$. If we wish a minimum 3-covering which starts with this 5-block we may continue with β or α, but we must then continue with $\alpha\beta$ or $\beta\alpha\beta$ respectively, thereby obtaining the blocks

$$(\alpha\beta\beta\alpha\alpha)\beta\alpha\beta, \qquad (\alpha\beta\beta\alpha\alpha)\alpha\beta\alpha\beta.$$

Further continuation subject to the condition that no 3-block repeats is impossible. Since $\theta(3) = 10$, there exists no transitive ray for which

$$\phi(n) = \theta(n) \qquad (n = 1, 2, 3).$$

It is conceivable that there might exist a transitive ray for which $\phi(n) = \theta(n)$ for all values of n exceeding some fixed integer m. That this is impossible is shown by the following previously unpublished theorem of R. Oldenburger.

THEOREM 11.1. *If the generating symbols reduce to two free symbols α and β there exists no transitive ray whose ergodic function $\phi(n)$ equals the covering index $\theta(n)$ for two successive values of $n > 1$.*

The proof as given by Oldenburger is essentially as follows.

We assume the theorem false. There then exists a minimum n-covering $H(n)$, $n > 2$, which starts with a minimum $(n-1)$-covering $H(n-1)$. No n-block repeats in $H(n)$, nor $(n-1)$-block in $H(n-1)$. We shall arrive at a contradiction.

Since α and β can be interchanged throughout it is no essential restriction to assume that in $H(n-1)$ the $(n-1)$-block β^{n-1} occurs before α^{n-1}. We shall then show that $H(n-1)$ has the form

$$H(n-1) = [\beta^{n-1}\alpha \cdot \cdot \cdot \beta\alpha^{n-1}], \tag{11.1}$$

and that $H(n)$ has the form

$$H(n) = H(n-1)[\alpha\beta \cdot \cdot \cdot \alpha\beta^{n}]. \tag{11.2}$$

The block of $H(n)$ which remains when $H(n-1)$ is removed will be denoted by J. To establish the correctness of the representations (11.1) and (11.2) one reasons as follows.

The block α^n appears in $H(n)$. It cannot appear wholly in $H(n-1)$, for this would imply repetition of α^{n-1} in $H(n-1)$, and thus $\beta\alpha^n$ appears in $H(n)$. Since β^{n-1} precedes α^{n-1} in $H(n-1)$, $\beta\alpha^{n-1}$ appears in $H(n-1)$. Since $\beta\alpha^{n-1}$ appears just once in $H(n)$, it must appear as the initial n-block of $\beta\alpha^n$, and since this latter block is not wholly in $H(n-1)$, $H(n-1)$ ends with $\beta\alpha^{n-1}$ and J begins with α. This α is of course followed by β.

The block β^n is in $H(n)$ and not wholly in $H(n-1)$. Since J begins with α, β^n must appear wholly in J, and hence $\alpha\beta^{n-1}$ and $\alpha\beta^n$ appear in J. It follows that β^{n-1}, which appears in $H(n-1)$, must appear at the left of $H(n-1)$; otherwise it would be preceded by α, and $\alpha\beta^{n-1}$ would appear twice in $H(n)$. Hence $H(n-1)$ begins with $\beta^{n-1}\alpha$. If J does not end with β^n, $\beta^n\alpha$ appears in J, and $\beta^{n-1}\alpha$ appears both in J and $H(n-1)$. Hence J ends with $\alpha\beta^n$ and the representations (11.1) and (11.2) are correct.

We now make use of the assumption that $n > 2$. Then $n-2 > 0$ and $\alpha\beta^{n-2}$ occurs in $H(n-1)$. This block cannot be followed by β without repetition of β^{n-1} in $H(n-1)$. Hence $\alpha\beta^{n-2}\alpha$ appears in $H(n-1)$. But $\beta^{n-2}\alpha$ already appears at the beginning of $H(n-1)$. Hence $H(n)$ cannot start with a minimum $(n-1)$-covering, and the proof of the theorem is complete.

We turn to the case where the only condition on the generating symbols is the restriction of inverses. This is the simplest case of strict dynamical significance. It arises in the case of surfaces of negative curvature with boundaries. Cf. Morse [1]. In this case one has a set of generating symbols of the form

$$a_1, \cdots, a_p; a_1^{-1}, \cdots, a_p^{-1} \qquad (p > 1), \tag{11.3}$$

with the restriction that successive symbols shall not be inverses. We shall prove the following theorem by suitably modifying Martin's proof.

THEOREM 11.2. *For generating symbols of the form* (11.3), *subject to no condition beyond the restriction of inverses, there exists a minimum r-covering in which no r-block repeats.*

The following rule for the formation of a minimum r-covering $H(r)$ was first given by Martin for the case of free generating symbols $c_1, \cdots, c_n$. Start with $c_1{}^{r-1}$ and continue successively adjoining the generating symbol of highest index consistent with the condition that no r-block be repeated. This will lead to a minimum r-covering in the case of free symbols c_i. For the case of inverses we adjoin the condition that successive symbols shall not be inverses. With this condition added the process will not always lead to a minimum r-covering. We are able however to prove the following lemma.

LEMMA 11.1. *A necessary and sufficient condition that the application of the Martin rule, subject to the restriction of inverses, shall lead to a minimum r-covering is that* c_1 *and* c_2 *shall not be inverses in the set* $c_1, \cdots, c_n$ *of generating symbols.*

Let M_r be the block of maximum length obtained by applying the Martin rule subject to the restriction of inverses. If A is any admissible $(r-1)$-block, the r-blocks

$$Ac_n, \cdots, Ac_1 \tag{11.4}$$

appear in M_r, if at all, in the order written. One of the blocks (11.4) is inadmissible since the last symbol in A cannot be followed by its inverse. If Ac_1 is in M_r, all other admissible blocks in (11.4) are in M_r. If Ac_1 is not admissible, Ac_2 is admissible and is in M_r only if all of its predecessors in (11.4) are in M_r. We set $E = c_1{}^{r-1}$. We continue with a proof of (a) and (b).

(a). *The block* M_r *ends with* E. If M_r ended with an $(r-1)$-block $A \neq E$, M_r could not begin with A, since M_r begins with E. A would then

appear in M_r at most $n-1$ times; otherwise some one of the $n-1$ admissible blocks c_iA would appear twice. Following A's last appearance we can continue M_r with c_2 (or c_1 if Ac_2 is inadmissible), contrary to the hypothesis that M_r ends with A. Hence M_r ends with E.

(b). *The block M_r ends with Ec_1.* Suppose M_r ends with an r-block $B \neq Ec_1$. Set $B = bE$ in accordance with (a). We are supposing that $b \neq c_1$. The block E appears n times in M_r; otherwise M_r could be continued. Hence Ec_1 appears in M_r but not at the end. But Ec_1 cannot be continued in accordance with the Martin rule. From this contradiction we infer the truth of (b).

We shall now make use of the assumption that c_1 and c_2 are not inverses and show that M_r contains all r-blocks.

We first observe that each admissible block Ec_j appears in M_r since M_r ends with Ec_1. Hence M_r contains n copies of E. It follows that M_r contains each admissible block c_jE. Let $B = b_1 \cdot \cdot \cdot b_r$ be an arbitrary admissible r-block. We shall show that M_r contains B.

Suppose that M_r does not contain B. Set $b_2 \cdot \cdot \cdot b_r = D$. Then $D \neq E$ by virtue of (b). Hence D appears at most $n-2$ times in M_r; otherwise each admissible block c_jD would appear in M_r, including B. Hence Dc_1 does not appear in M_r (or Dc_2, if Dc_1 is inadmissible). We now apply the same reasoning to Dc_1 (or Dc_2) that we have applied to B and infer that $b_3 \cdot \cdot \cdot b_rc_1c_1$ (or $b_3 \cdot \cdot \cdot b_rc_2c_1$) does not appear in M_r. Continuing we arrive at the conclusion that c_1^r does not appear in M_r, contrary to (b).

The condition of the lemma is accordingly sufficient.

To prove the condition necessary we suppose that $c_1 = c_2^{-1}$. If M_r were an r-covering c_2^r would appear in M_r, but not in the last position, by virtue of (a). But c_2^r cannot be continued without violating the Martin rule, or else the condition on inverses. Hence the condition is necessary, and the proof of the lemma is complete.

Theorem 11.2 is an immediate consequence of the lemma.

If there are n generating symbols subject merely to the restriction of inverses, there are $n(n-1)^{r-1}$ different admissible r-blocks so that the r-th covering index $\theta(n)$ has the value

$$\theta(r) = n(n-1)^{r-1} + r - 1.$$

With this understood we shall prove the following theorem.

THEOREM 11.3. *When there are n generating symbols subject merely to*

the restriction of inverses, each ergodic function $\psi(r)$ is at least $\theta(r)$ and there exists an ergodic function $\phi(r)$ such that

$$\phi(r) < \left(\frac{n-1}{n-2}\right) n(n-1)^{r-1} < \frac{n-1}{n-2}\,\theta(r). \tag{11.5}$$

Let H_r be a minimum r-covering and let H^*_r be the block obtained from H_r by omitting the first $r-1$ cymbols c_1. The ray

$$R = H_1 H^*_2 H^*_3 \cdots$$

is admissible and transitive. We note in particular that H^*_r is preceded by ${c_1}^{r-1}$ in R. For this ray $\phi(r)$ is at most the length of the block $H_1 H^*_2 \cdots H^*_r$, that is

$$\phi(r) \leqq \sum_{m=1}^{r} n(n-1)^{m-1}.$$

Hence

$$\phi(r) \leqq n\left[\frac{1-(n-1)^r}{1-(n-1)}\right] = \left(\frac{n-1}{n-2}\right) n\left[(n-1)^{r-1} - \frac{1}{n-1}\right],$$

and (11.5) follows at once.

We conclude this section by proving the following theorem.

THEOREM 11.4. *If $N(r)$ is the number of admissible r-blocks under our general conditions of admissibility, there exists an ergodic function $\phi(r)$ such that*

$$\phi(r) < (r+1)N(r), \tag{11.6}$$

while each ergodic function is at least $N(r) + r - 1$.

We begin by proving the following.

(a). *There exists an r-covering of length less than $(r+1)N(r)$.*

Statement (a) is true if $r = 1$. For if $a_1, \cdots, a_n$ *is* the set of generating symbols, there exists a 1-covering of the form $a_1 e_1 a_2 e_2 \cdots e_{n-1} a_n$ provided the symbols e_i are suitably chosen. This 1-covering has the length $2n-1 < 2N(1)$. We assume then that A_r is an r-covering of length less than $(r+1)N(r)$. We shall show how to form an $(r+1)$-covering of length less than $(r+2)N(r+1)$.

There are $N(r)$ distinct r-blocks in A_r and hence at least $N(r)-1$ distinct $(r+1)$-blocks. There remain at most

$$\mu = N(r+1) - N(r) + 1$$

$(r+1)$-blocks, say $B_1, \cdots, B_\mu$ not in A. The block

$$A_{r+1} = A_r e_1 B_1 e_2 B_2 e_3 \cdots e_\mu B_\mu,$$

in which the symbols e_i are chosen so as to make A_{r+1} admissible, is an $(r+1)$-covering. Let s_i be the length of A_i. We see that

$$s_{r+1} = s_r + (r+2)\mu.$$

By virtue of the hypothesis that $s_r < (r+1)N(r)$ we infer that

$$\begin{aligned} (11.7) \qquad s_{r+1} &< (r+1)N(r) + (r+2)[N(r+1) - N(r) + 1] \\ &= (r+2)N(r+1) - N(r) + r + 2. \end{aligned}$$

But for $r > 1$,

$$N(r) \geqq 2^r \geqq r+2,$$

so that (11.7) yields the relation

$$s_{r+1} < (r+2)N(r+1),$$

and the proof of (a) is complete.

In proving (a) we have defined r-coverings A_r, $r = 1, 2, \cdots$, such that A_r is an initial block of A_{r+1}. A ray whose first s_r symbols are those of A_r for each r will satisfy (11.6). The final statement of the theorem is immediate.

12. Category and sets of transitive rays. In this section we shall determine the category of the set of transitive rays or of sets of relatively transitive rays. More particularly, we shall determine the category of various subsets of transitive rays defined by limitations on the ergodic functions. In this way we obtain an idea of the distribution of transitive rays. A similar use can be made of measure theory (cf. § 13), but the development of this phase of the theory is deferred to a later paper.

In the category theory, the space used will be the space S of admissible rays with the Baire metric. Cf. § 4. Recall that the space S is compact, perfect, and totally disconnected. Let H be a perfect subset of S. A set A is of the *first category* relative to H if it is the sum of an at most enumerable set of nowhere dense (relative to H) subsets of H. These subsets will be called the components of A. The components of A may vary from the extreme

case where they are all null to the case where they are perfect. The complement $H - A$ of a set A of the first category is termed a *residual set* of the second category. Such a set is everywhere dense in H and has the power of the continuum. Cf. Hobson [1], p. 132.

Let R be a given ray. The rays obtained by deleting a finite initial block of R will be said to be *based* on R, those obtained by the inverse process of adding a finite block to R will be termed *extensions* of R. The rays which are limit rays in the space S of the set of rays based on R will be called *limit* rays of R. A ray R which is a limit ray of itself will be said to be *transitive relative to its limit rays*. It is thereby necessary and sufficient that R be non-periodic and that every block of R appear infinitely many times in R. The limit rays of a relatively transitive ray R form a closed set H. This set is identical with its derived set since each ray based on R is a limit ray of the set. The set H is accordingly perfect. Concerning the set H we have the following theorem.

THEOREM 12.1. *In the set H of limit rays of a relatively transitive ray R, the rays which are transitive relative to H form a residual set in H.*

Let $A_1, A_2, \cdots$ be an enumeration of the blocks of R. Let E_n be the subset of rays of H which contain no copy of A_n. The set E_n is closed and contains no rays based on R. It is nowhere dense in H since the rays based on R are everywhere dense in H. Finally the set S of rays which are transitive relative to H is of the form

$$S = H - \Sigma E_n,$$

and the theorem follows from the definition of a residual set in H.

The above sets E_n may all be null, in fact are all null if R is a subray of a non-periodic recurrent trajectory. In the case where R is transitive, that is transitive relative to the set of all admissible rays, the sets E_n are never null as the following lemma shows.

LEMMA 12.1. *The set K of rays which exclude an admissible r-block B for which $r > 2$ is closed, nowhere dense, and of the power of the continuum.*

It is clear that K is closed. That K is nowhere dense in S follows from the existence of a transitive ray R. For the rays based on R are everywhere dense in S and transitive. Since K is closed, it must then be nowhere dense.

It remains to prove that K is of the power of the continuum. To that end we consider the admissible rays

$$\alpha^{r_1}\beta\alpha^{r_2}\beta\alpha^{r_3}\beta\cdots, \qquad r_i = 2 \text{ or } 3, \tag{12.1}$$

formed from the free symbols α and β. This set is in one-to-one correspondence with the dyadic decimals

$$.r_1r_2r_3\cdots, \qquad r_i = 2 \text{ or } 3,$$

and hence is of the power of the continuum.

Similarly, the set

$$\beta^{r_1}\alpha\beta^{r_2}\alpha\beta^{r_3}\cdots, \qquad r_i = 2 \text{ or } 3, \tag{12.2}$$

consists of admissible rays and is of the power of the continuum.

A ray of the form (12.1) has no 3-blocks in common with a ray of the form (12.2). Since the length of B is at least 3, it cannot appear in a ray of (12.1) and a ray of (12.2). Hence at least one of the sets (12.1) and (12.2) is contained in K. Thus K must be of the power of the continuum.

The set of non-transitive rays is of the first category. But much more than this can be said, as the following theorem shows.

THEOREM 12.2. *The set Σ of non-transitive rays is everywhere dense in S and is the sum of countably many nowhere dense, closed sets, each of which is of the power of the continuum.*

That the set Σ is everywhere dense follows from the fact that an arbitrary block B and a sequence A of the symbols α can be combined in the form BeA to form a non-transitive ray. Rays of this form are everywhere dense in S. Let $A_1, A_2, \cdots$ be an enumeration of admissible r-blocks with $r > 2$. The set E_n of rays which exclude A_n is closed, nowhere dense and of the power of the continuum, in accordance with Lemma 12.1. The set Σ is the sum of the sets E_n, and the proof is complete.

THEOREM 12.3. *If $f(n)$ is an arbitrary positive function of the positive integers n, the set F of transitive rays with ergodic function $\phi(n) > f(n)$, for all values of n, has the power of the continuum.*

Let $A_1, A_2, \cdots$ be an enumeration of admissible blocks consistent with their lengths. Let $r_1, r_2, \cdots$ be an arbitrary sequence of positive integers and let $c_1c_2\cdots$ be an arbitrary sequence of the symbols α and β. We introduce the block

$$B_i = \alpha^{r_i}c_i\alpha^{r_i}e_iA_i \qquad (i = 1, 2, \cdots),$$

choosing e_i so that B_i is admissible. We then introduce the ray

$$E = B_1 b_1 B_2 b_2 \cdots,$$

choosing the symbols b_i so that the ray E is admissible. It is clear that E is transitive.

Let $\phi(n)$ be the ergodic function of E. We choose r_1 so that $\phi(n) > f(n)$ for $n = 1, 2, 3$. Let B be the initial $f(4)$-block of E. If r_2 is chosen sufficiently large, blocks $A_1, A_2, \cdots, A_p$ entering into the definition of B will all be 1-blocks. In fact, if r_2 is sufficiently large only A_1 will enter into the definition of B. If A_i is a 1-block, $e_i A_i b_i$ is a 3-block so that B cannot contain β^4. With r_2 so chosen, $\phi(4) > f(4)$.

Proceeding inductively with $m > 2$ we choose r_m so large that in the initial $f(m+2)$-block C of E, at most $(m-1)$-blocks A_i are employed. The corresponding blocks $e_i A_i b_i$ are at most $(m+1)$-blocks so that β^{m+2} does not appear in C. Hence

$$\phi(m+2) > f(m+2) \qquad (m = 3, 4, \cdots). \tag{12.3}$$

Thus $\phi(n) > f(n)$ for all values of n.

The above symbols c_i can be chosen arbitrarily from the set α, β, the r_i's remaining fixed. For fixed r_i's there are then as many different rays E as there are different sequences $c_1 c_2 \cdots$, that is a number equal to the power of the continuum.

The proof of the theorem is complete.

The set F of the theorem is not closed. The set of transitive rays not in F is open in the set of all transitive rays so that F is everywhere dense only if it includes all transitive rays.

THEOREM 12.4. *Given an arbitrary function $f(n)$, the set ω of transitive rays R with ergodic functions $\phi(n) \leqq f(n)$ for n exceeding some integer dependent on R is a set of the first category.*

Let D_m be the set of transitive rays for which $\phi(n) \leqq f(n)$ for n exceeding m. The set D_m is closed. But we have seen in Theorem 12.2 that the set of intransitive rays is everywhere dense. Hence D_m is nowhere dense. But $\omega = \Sigma D_m$, and the proof is complete.

Let $H(n)$ be a block of minimum length $\theta(n)$ covering all admissible n-blocks. We have termed $\theta(n)$ the n-th covering index. As we have seen in § 11, there is in general no transitive ray with ergodic function $\phi(n) = \theta(n)$. We have however the following theorem.

THEOREM 12.5. *The set G of transitive rays with ergodic function $\phi(n)$ such that*

$$\liminf_{n\to\infty} \frac{\phi(n)}{\theta(n)} = 1, \tag{12.4}$$

where $\theta(n)$ is the n-th covering index, is a residual set of the second category.

We begin with a proof of the following.

(a). *The set G is not empty.*

To prove (a) let δ_n be a sequence of positive constants converging to zero. We shall consider a ray of the form

$$R = H(r_0)e_1H(r_1)e_2H(r_2)\cdots,$$

where $H(n)$ is defined as above and the symbols e_i are chosen so that R is admissible. The integers r_i are chosen as follows. Let $r_0 = 1$. Proceeding inductively we suppose that r_{n-1} is defined and that the block

$$H(r_0)e_1H(r_1)e_2\cdots e_{n-1}H(r_{n-1})e_n$$

has the length m_n. We then choose r_n so large that

$$\frac{\theta(r_n) + m_n}{\theta(r_n)} < 1 + \delta_n. \tag{12.5}$$

If $\phi(n)$ is the ergodic function of the resulting ray R, it is clear that

$$\phi(r_n) \leqq \theta(r_n) + m_n,$$

since the initial block of length $\theta(r_n) + m_n$ in R contains $H(r_n)$ and hence all r_n-blocks. The function $\phi(n)$ will satisfy (12.4) by virtue of (12.5), and the proof of (a) is complete.

If A is an arbitrary admissible r-block, the ray $X = AeR$ is admissible if e is suitably chosen. Moreover the ray R and the ray X have ergodic functions ϕ and ψ respectively such that

$$\theta(n) \leqq \psi(n) \leqq \phi(n) + r + 1.$$

Hence X satisfies the condition (12.4) and is in G. We conclude that the rays of G are everywhere dense.

To come to a proof of the theorem let m be a positive integer and p a rational number > 1. The rays for which the transitivity index $\psi(m)$ satisfies the condition

$$\psi(m) > \theta(m)p$$

form a closed set $D_m{}^p$. The intersection

$$E_q{}^p = D_q{}^p D^p{}_{q+1} D^p{}_{q+2} \cdots$$

is closed. The complement of G is the sum of the sets $E_q{}^p$ as q ranges over all positive integers and p ranges over all rational numbers > 1. But G is everywhere dense so that $E_q{}^p$ is nowhere dense, and the proof of the theorem is complete.

THEOREM 12. 6. *The set Z of transitive rays with ergodic functions $\phi(n)$ such that*

$$\liminf_{n\to\infty} \frac{\phi(n)}{\theta(n)} > 1$$

is everywhere dense, has the power of the continuum, and is of the first category.

The set Z includes the set of transitive rays for which $\phi(n) > 2\theta(n)$, and so has the power of the continuum in accordance with Theorem 12. 3. The set Z is a subset of the complement of the residual set G of Theorem 12. 5 and so is of the first category. It remains to show that Z is everywhere dense.

If R is a ray of Z, any ray R' based on R is again transitive. If the initial symbol of R' is the $(r+1)$-st symbol of R, the ergodic functions $\psi(n)$ and $\phi(n)$ of R' and R, respectively, satisfy the conditions

$$\psi(n) \geqq \phi(n) - r, \qquad \liminf_{n\to\infty} \frac{\psi(n)}{\theta(n)} > 1.$$

Hence R' belongs to Z. But the rays R' based on the transitive ray R are everywhere dense so that Z is everywhere dense as stated.

The proof of the theorem is complete.

13. Trajectories with the separation property. The results of this section will form the basis of a more extensive treatment to be presented later. No proofs will be given in this place.

Let t be a trajectory involving two generating symbols α and β. By the α-length (β-length) of a block B of t we shall mean the number of α's (β's) in B. The trajectory t will be said to have the *separation property* and be of *type S* if the following holds.

The α-lengths of any two n-blocks of t differ by at most 1.

The preceding condition implies the corresponding condition for β-lengths. Trajectories of type S arise in many ways. In particular let

$$\frac{d^2y}{dx^2} + f(x)y = 0 \tag{13.1}$$

be a differential equation in which $f(x)$ is continuous and has the period 1. Let $u(x)$ be a solution of (13.1) which vanishes at least once, but not identically. We are concerned with the distribution of the zeros of $u(x)$. Without loss of generality we can suppose that none of these zeros occur at the points $x = n$, where n is an integer, positive, negative or zero. For the zeros of $u(x)$ are countable, and for our purposes any transformation of the form $\bar{x} = x + c$ is admissible. With this understood let a trajectory t be defined as follows. Let each point on the x-axis at which $u(x) = 0$ be replaced by the symbol α, and each point $x = n$ be replaced by the symbol β. The resulting trajectory t will be of type S. With trivial exceptions every trajectory of type S can be obtained in this way. The study of trajectories of type S is thus a study of the zeros of solutions of equations of the form (13.1).

Let $\alpha_n(\beta_n)$ denote the α-length (β-length) of some n-block of a trajectory t of type S. It can be shown that there exists a number θ, $0 \leqq \theta \leqq 1$, such that

$$\lim_{n\to\infty} \frac{\alpha_n}{n} = \theta, \qquad \lim_{n\to\infty} \frac{\beta_n}{n} = 1 - \theta.$$

The number θ will be termed the *rotation number* of t. Cf. Poincaré [1].

The trajectory is not uniquely determined by the rotation number. If θ is irrational, there are infinitely many trajectories of type S having θ as rotation number and these trajectories are all recurrent. Thus trajectories of type S are in general recurrent. A simple exceptional example is the trajectory in which all the symbols are α except for one β, in which case the rotation number is 1.

Given θ, $0 \leqq \theta \leqq 1$, the recurrent trajectories of type S which have θ as rotation number form a minimal set which consists of a single periodic trajectory if and only if θ is rational. Since all of the trajectories of one of these minimal sets have the same recurrency function, it can be denoted by $R_\theta(n)$. The behavior of $R_\theta(n)$ varies widely with the choice of θ, but in all cases

$$\liminf_{n\to\infty} \frac{R_\theta(n)}{n} \leqq 4.$$

There exist irrational values of θ, and thus non-periodic recurrent trajectories such that

$$\limsup_{n\to\infty} \frac{R_\theta(n)}{n} \leqq 6.5.$$

On the other hand, given an arbitrary function $f(n)$, there exist irrational values of θ such that $R_\theta(n) > f(n)$ for infinitely many values of n.

If certain exceptional θ-sets of (Lebesgue) measure zero are excluded, $R_\theta(n)$ displays great regularity in its asymptotic behavior. We can state that, except for a θ-set of measure zero,

$$\limsup_{n\to\infty} \frac{R_\theta(n)}{n} = +\infty,$$

while if $c > 1$, except for a θ-set of measure zero,

$$\limsup_{n\to\infty} \frac{R_\theta(n)}{n^c} \leqq 1.$$

14. Typical open questions. Let R be a ray of the form $c_0c_1c_2\cdots$ and let t be a trajectory containing R as a subray. Among the positive limit trajectories of t are found one or more distinct minimal sets of trajectories. Cf. Theorem 6.1. Suppose v_1 of these sets are periodic trajectories, and v_2 of these sets general minimal sets. The number v_1 is at most aleph null, and v_2 is at most aleph. That v_2 may equal aleph will be shown by an example in a later paper. The numbers v_1 and v_2 are uniquely determined by R and will be called the *limit indices* of R of the first and second types respectively. Our first question is as follows.

(1). What are the limit indices of algebraic numbers, regarding decimal representations as rays? Are these rays ever transitive or recurrent? What are the answers to similar questions concerning π and e?

(2). Let t be a non-periodic recurrent trajectory. In all known examples $R(n)/n$ fails to converge. Is this true in general?

(3). If t is a non-periodic recurrent trajectory, $R(n) \geqq kn$ where $k \geqq 2$. Cf. Theorem 7.5. What is the least value of k such that $R(n) \geqq kn$ for each non-periodic recurrent trajectory?

(4). In the case of general admissibility certain blocks and finite sequences of blocks are prohibited. Knowing these prohibitions, is it possible to give a constructive rule for the formation of a minimum block covering all admissible r-blocks, and what is the most general such r-covering?

(5). Does there exist an ergodic function $\phi(n)$ such that the ratio of $\phi(n)$ to $\theta(n)$ tends to the limit 1? Cf. Theorem 12.5.

(6). Is a theorem similar to that of Oldenburger true in case the generating symbols suffer the restriction of inverses, or in the general case?

(7). What is the most general minimal set with a given recurrency function $R(n)$? With permutation index $P(n)$? With $P(n)$ and $R(n)$ both given? These questions should be answered first when there are just two generating symbols.

These questions seem to point to an algebraic analysis. Questions more intimately connected with measure theory, or the analysis of space forms will be presented in a later paper.

THE INSTITUTE FOR ADVANCED STUDY,
AND
BRYN MAWR COLLEGE.

BIBLIOGRAPHY.

ARTIN, E.

1. "Ein mechanisches System mit quasiergodischen Bahnen," *Abhandlungen aus dem Mathematischen Seminar der Hamburgischen Universität*, Bd. 3 (1924), pp. 170-175.

BESICOVITCH, A. S.

1. *Almost periodic functions*, Cambridge, University Press (1932).

BIRKHOFF, G. D.

1. "Quelques théorèmes sur le mouvement des systèmes dynamiques," *Bulletin de la Société Mathématique de France*, t. 40 (1912), pp. 303-323.
2. "Dynamical systems," *American Mathematical Society Colloquium Publications*, vol. 9 (New York, 1927).
3. "Nouvelles recherches sur les systèmes dynamiques," *Memoriae Pont. Acad. Scient. Novi Lyncaei* (Série 3), t. 1, pp. 85-216.
4. "Sur le problème restreint des trois corps" (Second Mémoire), *Annali della R. Scuola Normale Superiore di Pisa* (Serie II), t. 5 (1936), pp. 1-42.

HADAMARD, J.

1. "Les surfaces à courbure opposées et leur lignes géodésiques," *Journal de Mathématiques pures et appliquées* (5th series), t. 4 (1898), pp. 27-73.

HAUSDORFF, F.

1. *Mengenlehre*, Leipzig, Walter de Gruyter (1927).

HEDLUND, G. A.

1. "On the metrical transitivity of the geodesics on closed surfaces of constant negative curvature," *Annals of Mathematics*, vol. 35 (1934), pp. 787-808.
2. "A metrically transitive group defined by the modular group," *American Journal of Mathematics*, vol. 57 (1935), pp. 668-678.
3. "Two-dimensional manifolds and transitivity," *Annals of Mathematics*, vol. 37 (1936), pp. 534-542.

HILBERT, D.

1. "Über Flächen von konstanter Gaussscher Krümmung," *Transactions of the American Mathematical Society*, vol. 2 (1901), pp. 87-99.

HOBSON, E. W.

1. *The Theory of Functions of a Real Variable*, Cambridge, University Press (1921).

KOEBE, P.

1. "Riemannsche Mannigfaltigkeiten und nicht euklidische Raumformen," *Sitzungsberichte der Preussischen Akademie der Wissenschaften* (1927), pp. 164-196; (1928), pp. 345-442; (1929), pp. 414-457; (1930), pp. 304-364, 505-541; (1931), pp. 506-534.

MARTIN, M. H.

1. "A problem in arrangements," *Bulletin of the American Mathematical Society*, vol. 40 (1934), pp. 859-864.
2. "The ergodic function of Birkhoff," *Duke Mathematical Journal*, vol. 3 (1937), pp. 248-278.

MORSE, M.

1. "A one-to-one representation of geodesics on a surface of negative curvature," *American Journal of Mathematics*, vol. 43 (1921), pp. 33-51.
2. "Recurrent geodesics on a surface of negative curvature," *Transactions of the American Mathematical Society*, vol. 22 (1921), pp. 84-100.
3. "A fundamental class of geodesics on any closed surface of genus greater than one," *Transactions of the American Mathematical Society*, vol. 26 (1924), pp. 25-61.
4. "Instability and transitivity," *Journal de Mathématiques pures et appliquées* (9th series), t. 14 (1935), pp. 49-71.

MYRBERG, P. J.

1. "Einige Anwendungen der Kettenbrüche in der Theorie der binären quadratischen Formen und der elliptischen Modulfunktionen," *Annales Acad. Sc. Fennicae*, t. 23 (1924).
2. "Ein Approximationssatz für die Fuchsschen Gruppen," *Acta Mathematica*, t. 57 (1931), pp. 389-409.

NIELSEN, J.

1. "Om geodaetiske Linier i lukkede Mangfoldigheder med konstant negativ Krumning," *Matematisk Tidskrift* B, (1925), pp. 37-44.
2. "Untersuchungen zur Topologie der geschlossenen zweiseitigen Flächen," *Acta Mathematica*, t. 50 (1927), pp. 189-358.

POINCARÉ, H.

1. "Sur les courbes définies par les équations différentielles," *Journal de Mathématiques pures et appliquées* (Série 4), t. 1 (1885), pp. 167-244.

ROBBINS, H. E.

1. "On a class of recurrent sequences," *Bulletin of the American Mathematical Society*, vol. 43 (1937), pp. 413-417..

PERIODIC ORBITS IN THE RESTRICTED PROBLEM OF THREE BODIES IN THEIR RELATION TO HILL'S WORK ON THE MOTION OF THE MOON.

By Elis Strömgren.

The first volume of the *American Journal of Mathematics* (1878) contains two papers by G. W. Hill that have become of fundamental importance in the theory of the motion of the moon, and which, besides, contain a few remarks that may well be looked upon as the starting point of important chapters of that special case of the problem of three bodies known as "The restricted problem of three bodies," or " Problème restreint."

The Problème restreint was first defined by Jacobi: Three bodies are given, two of which have finite masses and are moving around each other in circular orbits according to the laws of the problem of two bodies. The problem is to study the motion of the third body, assumed to have an infinitely small mass, when moving in the plane of the two finite masses under the influence of the attraction from these two masses. With regard to the mass ratio, it should be noticed that Poincaré in his investigations of the restricted problem always assumes that one of the two finite masses is very small relative to the other, thus reducing the problem to a perturbational problem, which can be solved with the aid of series expressions in powers of a small quantity, but which on the other hand enables us to study the nature of the motion in the neighbourhood of known exact solutions only.

In the work carried out at the Copenhagen Observatory on the other hand, the two finite masses have been assumed to be equal, and the scope of the problem is essentially enlarged, it being desired to follow all investigated classes of orbits throughout the whole of their evolution. This problem, however, not being a simple perturbational problem, could only be solved with the aid of *numerical* methods of integration.

The cusps in Hill's variational curves. Plate III in vol. 1 of the *American Journal of Mathematics* gives (cf. Fig. 1 in this paper) a series of lunar orbits computed by Hill in the idealized variational problem, in which the parallactic factor is put equal to zero. In the diagram we see four variational curves from 1 to 4, of which the last has two symmetrical cusps, one above, and one below.

With regard to the orbits in a continued evolution of these variational curves Hill says as follows on p. 259 (*Collected Works*, I, 334): "It is certain that the orbits, if they do exist, do not intersect the line of quadratures, and that the moons describing them would make oscillations to and fro, never departing as much as 90° from the point of conjunction or of opposition." This was only a conjecture, and it was shown to be false by Poincaré in *Méthodes Nouvelles*, I, 106 ff. (1892). Here Poincaré shows by a simple mathematical procedure that the cusps in Hill's variational curves in the further evolution simply become *loops*. Later (1907) in the *Astronomische*

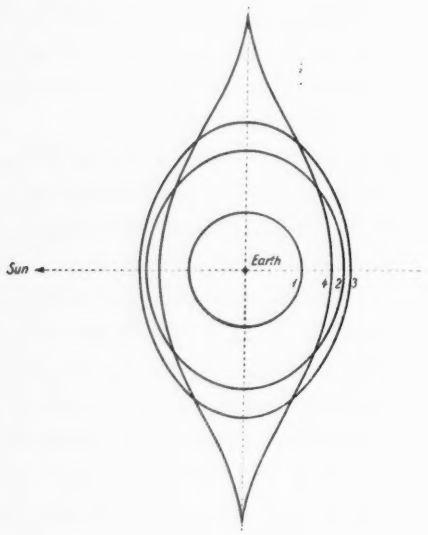


Fig. 1.
Hill's variational curves.

Nachrichten No. 4155, the author of this article drew attention to the fact that, in the restricted problem, loops and cusps can exist at any point in the plane, with the exception of certain special points. With regard to these loops and cusps, the author found a general theorem, to which we shall return.

Hill's division of the plane of motion into permitted and forbidden regions. In the first of the two papers in the first volume of the *American Journal of Mathematics* quoted above, Hill, with the aid of the so-called Jacobian integral, makes his famous investigation of the permitted and the

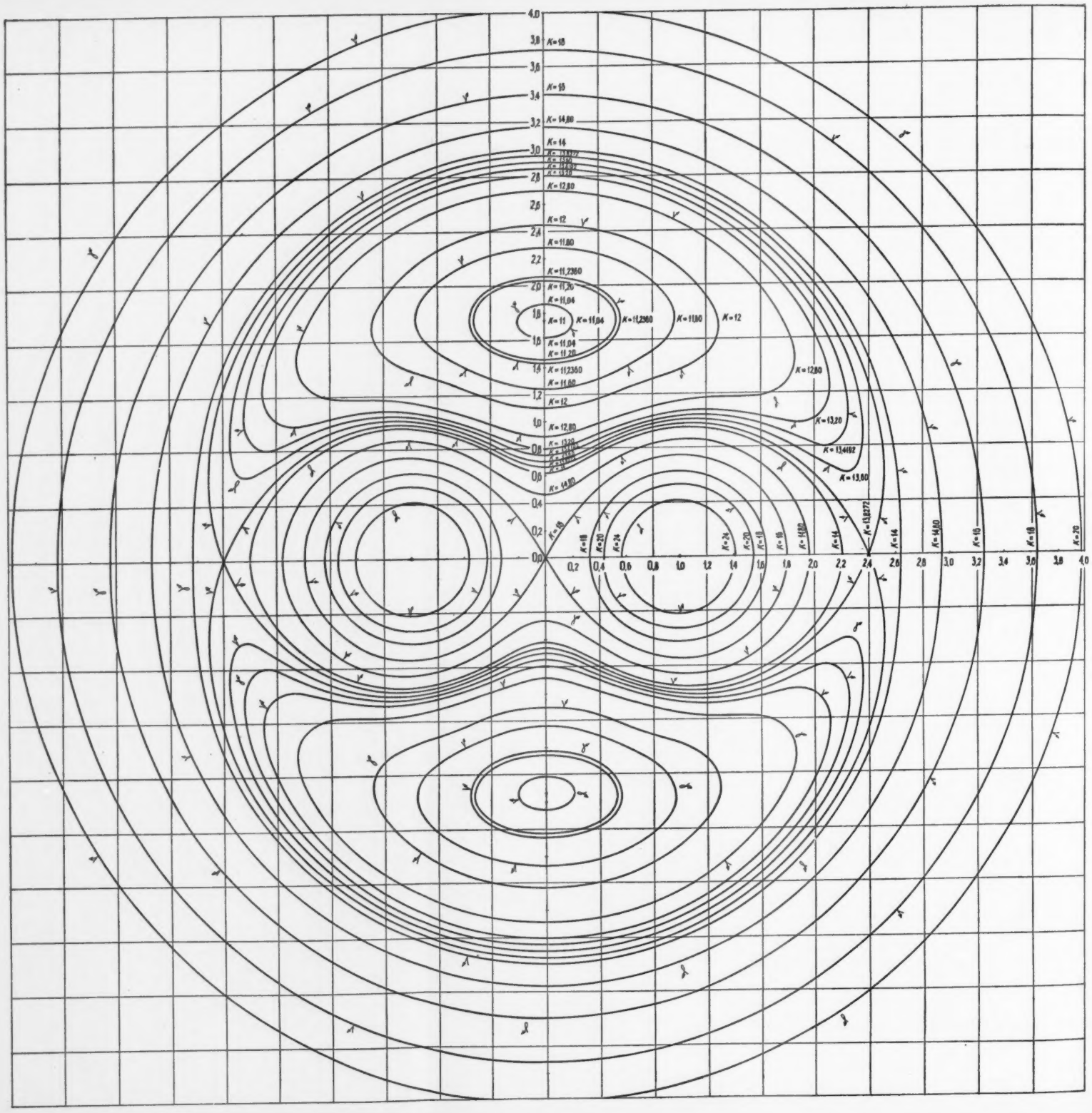


PLATE I.

Zero velocity-curves in the restricted problem of three bodies, with $m_1 = m_2$.

forbidden regions in the plane of motion. The division is made with the aid of the so-called curves of zero velocity. These curves are shown in Fig. 2 for the mass ratio of the two finite masses $m_1 = 10m_2$ used by G. H. Darwin in his "Periodic orbits." Plate I shows the curves for the mass ratio $m_1 = m_2$

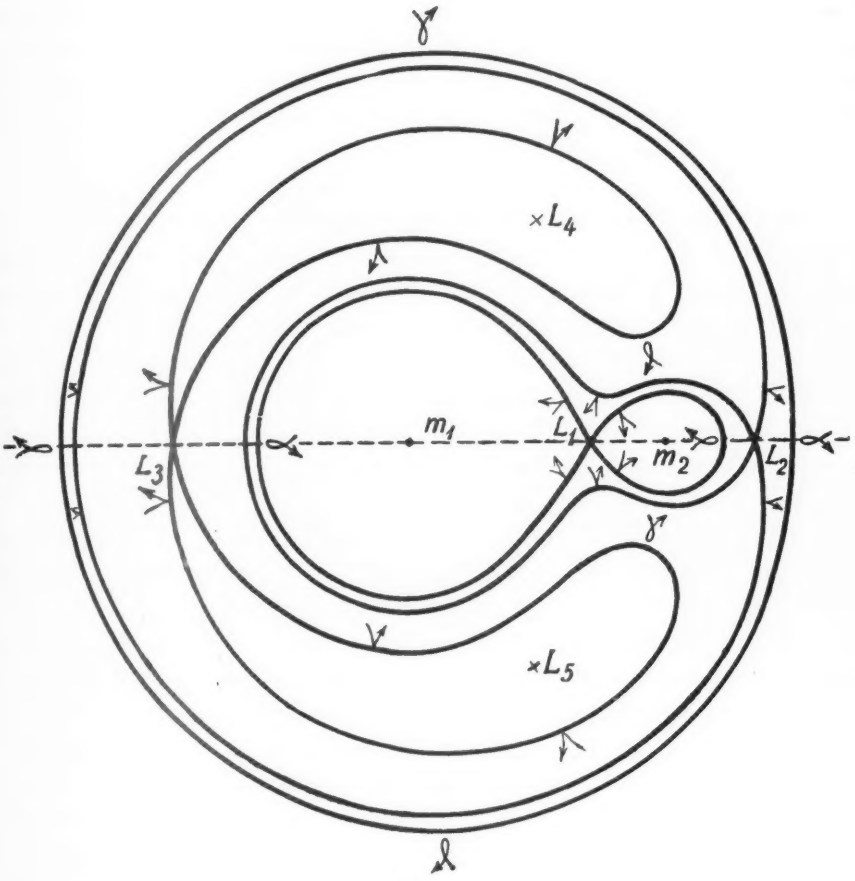


Fig. 2.

Zero velocity-curves in the restricted problem of three bodies, with $m_1 = 10m_2$.

suggested by Thiele and adopted in the work at the Copenhagen Observatory (Plate I is based upon a drawing made by the Moscow astronomer N. Moisseiev. The scale and the values of the so-called Jacobian constant are the same as in the work at the Copenhagen Observatory, cf. for instance Publ. No. 100 of the Copenhagen Observatory, or *Bulletin Astronomique*, Deuxième Série, Tome IX, Fasc. 2, 1935). In both of these drawings we have had cusps and loops drawn

especially, to illustrate the general theorem referred to above (*Astronomische Nachrichten* No. 4155): *When, located on that part of the normal of a zero velocity curve which belongs to the permitted region, we look in the direction toward the curve, then the motion of the infinitely small mass in a cusp is always from right to left.* As is seen from Fig. 2 and Plate I the direction of motion in a loop also follows immediately from this theorem. It should be noticed, that in these diagrams, as in Fig. 1 and all the following diagrams, the X-axis is the line from m_1 to m_2, which, in a fixed coördinate system is rotating with uniform velocity. The 5 points of libration are denoted by L_1, L_2, L_3, L_4, L_5 in Fig. 2.

The evolution of a cusp during the evolution of a class of orbits. The programme of the work at the Copenhagen Observatory originally was this: to find all classes of simply-periodic orbits around the finite masses and around the points of libration, and to follow these classes through the whole of their evolution. This programme has at the present time, with one single exception (cf. Publ. 100, pp. 26-27), been completely carried out. As the work progressed, however, this programme was automatically considerably extended, as we were, during the investigations, led to study quite different possibilities of finding periodic solutions in the problem.

Among the classes of orbits on our original programme was, amongst others, the following: It was clear a priori that at great distances from the origin orbits with slow, almost circular motion would exist, which in a fixed coördinate system were *direct* (motion in the same direction as the two finite masses), and—on account of the rotation of the m_1m_2-axis—rapid *retrograde* in the rotating coördinate system. Fig. 3 shows a number of orbits in this class. Far out we have an almost circular orbit ($K = 14.4361$). Then we see an orbit ($K = 12.60816$) of markedly oval shape, and still further inwards we encounter an orbit ($K = 11.76718$) with a marked indentation. This indentation develops in the evolution that follows into a cusp, and then into a loop.

Fig. 4-7 show how the evolution is continued. The evolution ends in the point of libration L_4 *with asymptotic motion into as well as out from this point of libration, on the ingoing branch with an infinite number of steadily decreasing windings, with the velocity converging toward zero,* on the outgoing branch the reverse. Fig. 8 shows the ingoing branch only.

This was an example of the asymptotic motions predicted by Poincaré in *Méthodes Nouvelles* on a theoretical basis, the asymptotic motion being here periodic also.

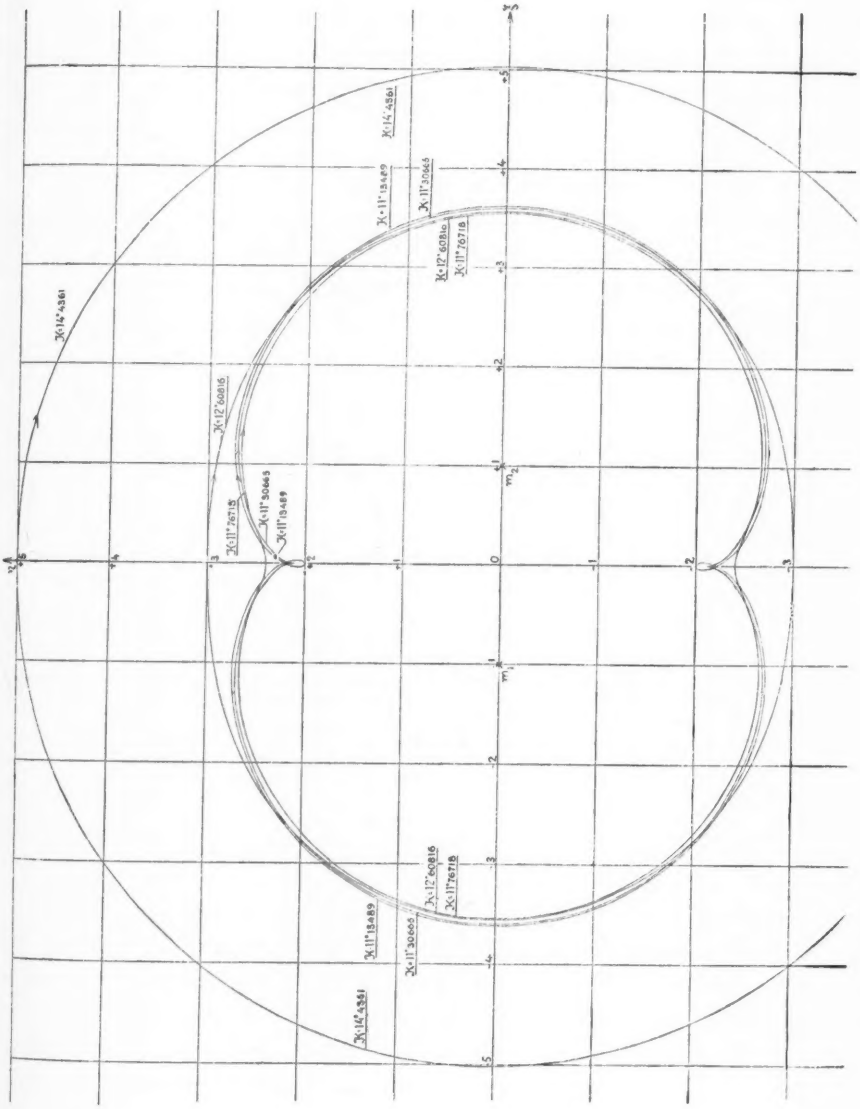


Fig. 3.
A class of periodic orbits in the restricted problem of three bodies.

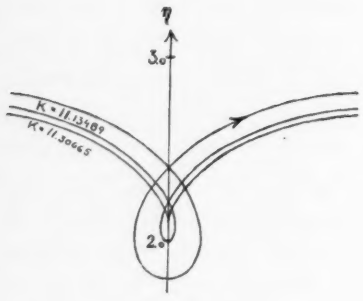


Fig. 4.

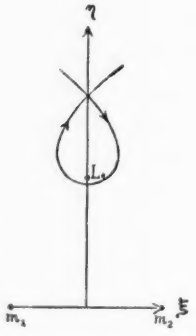


Fig. 5.

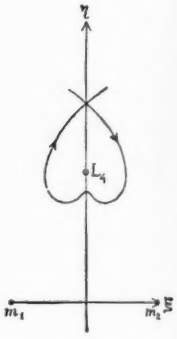


Fig. 6.

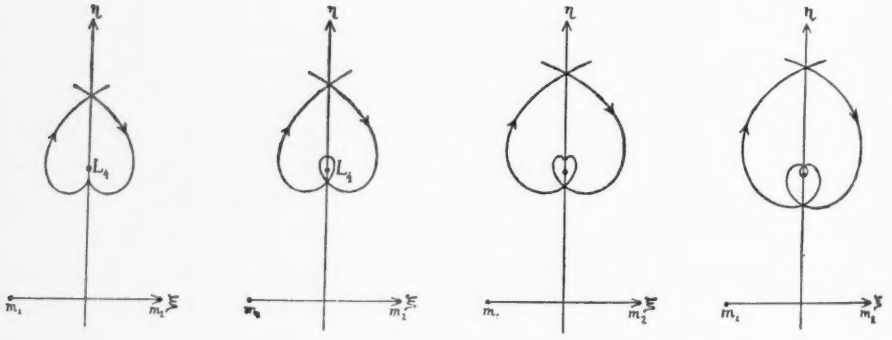


Fig. 7.

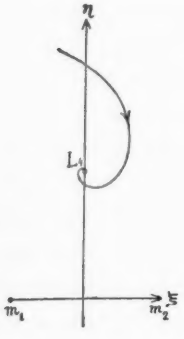


Fig. 8.

The evolution of cusps and loops up to asymptotic motion.

Asymptotic orbits, that are also periodic, with an infinitely long period. We have seen how the class of orbits shown in Fig. 3, which begins at infinity, ends with a periodic orbit of infinite period, asymptotic to the two points of libration L_4 and L_5. This is a special case of a general theorem, that was deduced in an empirical way as a result of our investigations: It has been found that, as far as our experience goes, a class of periodic orbits always, either

I) goes back into itself, so as to form a class closed in itself,

or II) has a natural beginning and a natural ending. Such a natural beginning or natural ending always takes place:

either a) in one of the two finite masses, or in both,

or b) in one of the 5 libration points, or in two of them,

or c) at infinity.

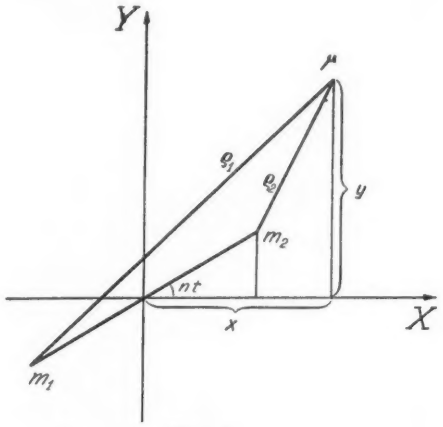


Fig. 9.

It now turned out that the study of the class of orbits shown in Fig. 3 had very extensive consequences. We see how an asymptotic-periodic orbit acts as the boundary orbit of a class of periodic orbits. The following idea was now rather obvious: If we can find other asymptotic orbits, which are also periodic (with infinitely long period), will it not then turn out that these orbits are boundary orbits of classes of periodic orbits?

The way that had to be followed now was very simple. We computed various orbits, that were asymptotic from the point of libration L_4.

We started the calculations with the aid of the theoretical expressions for the motion infinitely close to the libration point.

Fig. 9 defines the problem in the fixed coördinate system with center of

gravity of m_1 and m_2 at the origin. The differential equations of the motion of the infinitely small mass are, in the fixed coördinate system, and with our choice of units, as follows:

$$(1)\qquad \left\{\begin{aligned} \frac{d^2x}{dt^2} &= -4\left\{\frac{x+\cos t}{\rho_1^{\,3}}+\frac{x-\cos t}{\rho_2^{\,3}}\right\}\\ \frac{d^2y}{dt^2} &= -4\left\{\frac{y+\sin t}{\rho_1^{\,3}}+\frac{y-\sin t}{\rho_2^{\,3}}\right\}\\ \rho_1^{\,2} &= (x+\cos t)^2+(y+\sin t)^2\\ \rho_2^{\,2} &= (x-\cos t)^2+(y-\sin t)^2. \end{aligned}\right.$$

Referred to the coördinate system rotating with the line m_1m_2 the equations are:

$$(2)\qquad \left\{\begin{aligned} \frac{d^2\xi}{dt^2}-2\frac{d\eta}{dt}-\xi &= -4\left\{\frac{\xi+1}{\rho_1^{\,3}}+\frac{\xi-1}{\rho_2^{\,3}}\right\}\\ \frac{d^2\eta}{dt^2}+2\frac{d\xi}{dt}-\eta &= -4\left\{\frac{\eta}{\rho_1^{\,3}}+\frac{\eta}{\rho_2^{\,3}}\right\}\\ \rho_1^{\,2} &= (\xi+1)^2+\eta^2\\ \rho_2^{\,2} &= (\xi-1)^2+\eta^2. \end{aligned}\right.$$

If we transfer the origin of the coördinate system to L_4 and neglect powers of the coördinates higher than the first we obtain the following equations of motion:

$$(3)\qquad \left\{\begin{aligned} \frac{d^2x}{dt^2}-2\frac{dy}{dt} &= \frac{3}{4}x\\ \frac{d^2y}{dt^2}+2\frac{dx}{dt} &= \frac{9}{4}y. \end{aligned}\right.$$

which are, therefore, the differential equations of the infinitesimal motion near L_4. The solution for the ingoing motion is given in terms of the functions

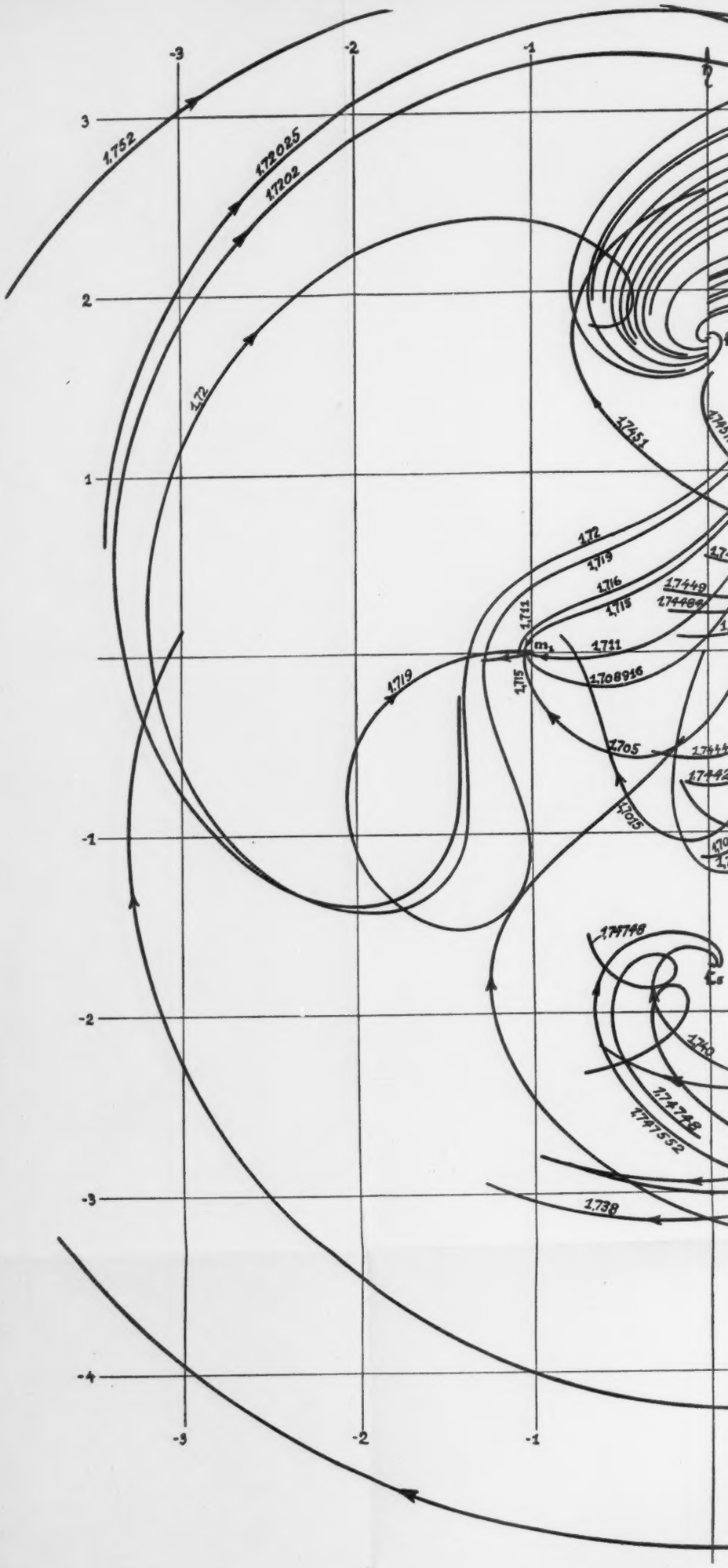
$$e^{-at}\sin bt \quad\text{and}\quad e^{-at}\cos bt,$$

that for the outgoing motion in terms of

$$e^{+at}\sin bt \quad\text{and}\quad e^{+at}\cos bt,$$

i. e. we have an asymptotic motion.

We follow the motion with the aid of the infinitesimal equations up to a point, which is very close to L_4, and then continue the solution with the aid of numerical integration of the exact differential equations (2).



Asymptotic o

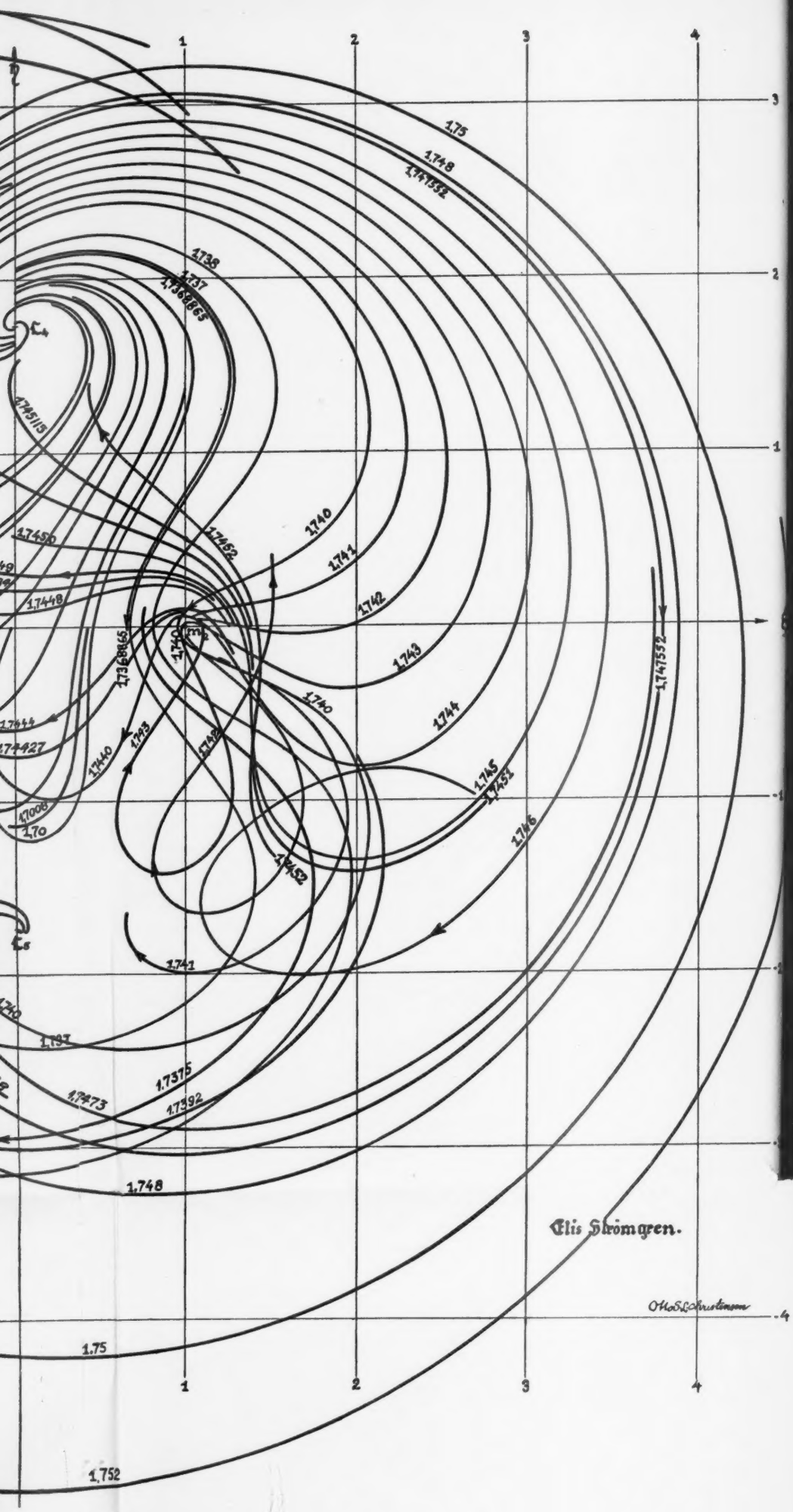


PLATE II.

tic orbits from the libration point L_4.

UNIV. OF MICH.

If we follow the motion until the orbit cuts the X-axis for the first time, we shall, in case the orbit should be perpendicular to the X-axis, have a periodic orbit–due to the symmetry of the differential equations. The orbit will continue symmetrically with regard to the X-axis, and asymptotically approach L_5. We may imagine the orbit continued in an orbit symmetrical with respect to the Y-axis, back to L_4, where the motion started.

The problem of finding such orbits, which are perpendicular to the X-axis, presents no difficulties. The method is to compute a number of orbits that cut the X-axis under various angles, interpolating finally between such of these orbits as make an angle of 90° with the X-axis.

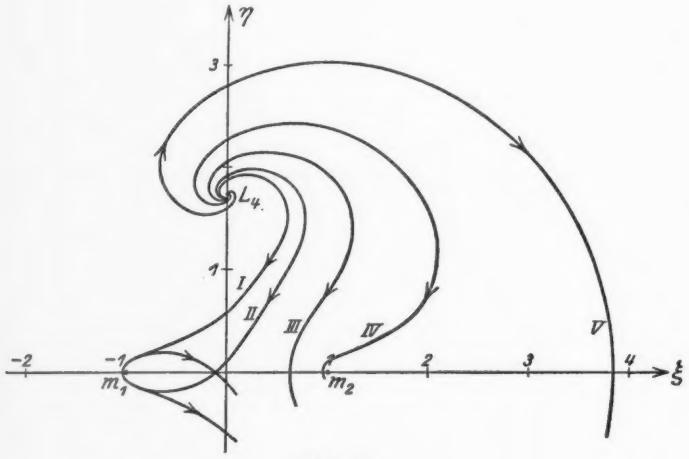


Fig. 10.

The 5 asymptotic-periodic orbits which are already perpendicular to the x-axis after a quarter of a revolution.

We have in Copenhagen made a large number of such calculations. Plate II gives the essential of the results obtained. Examining this drawing we find that 5 orbits exist, not more and not less, which are already perpendicular to the X-axis after a quarter of a revolution. Fig. 10 shows these five asymptotic-periodic orbits (for the sake of simplicity only about half of the orbits have been reproduced).

The question now arises, what does it mean that, in this way, a number of orbits exist, that are periodic (with infinitely long periods) and at the same time asymptotic to and from L_4 and L_5? Now, if for instance we investigate

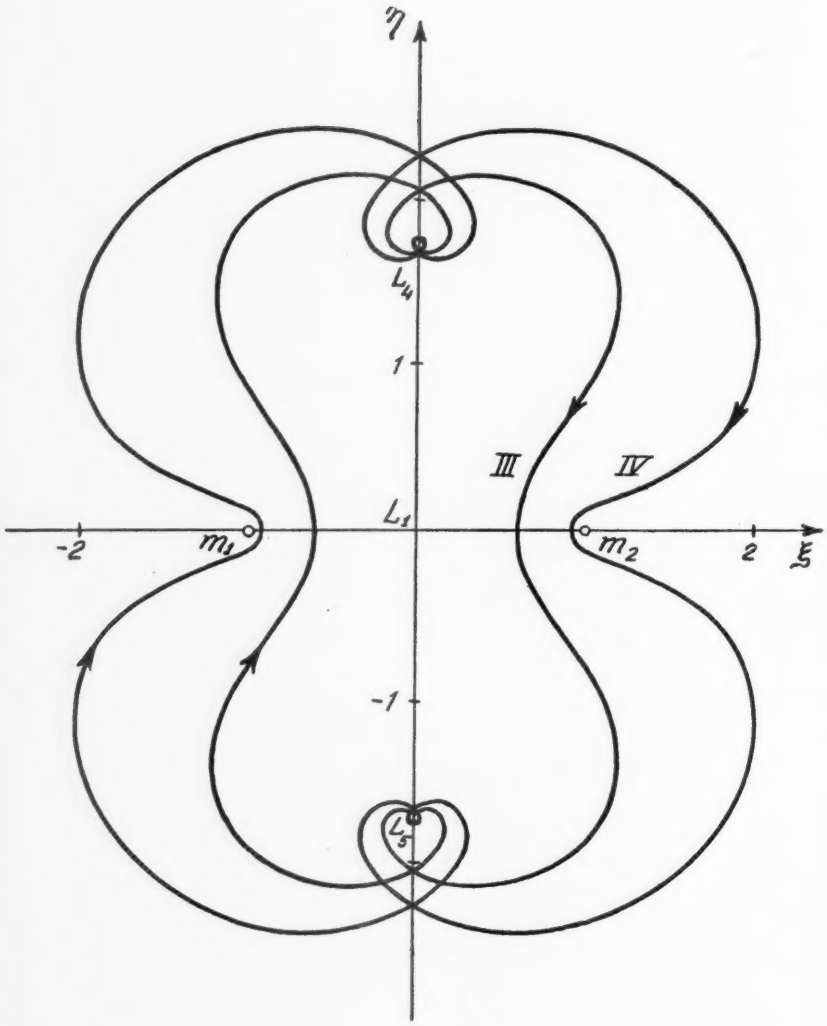


Fig. 11.

Two asymptotic orbits as boundary orbits of a class of periodic orbits.

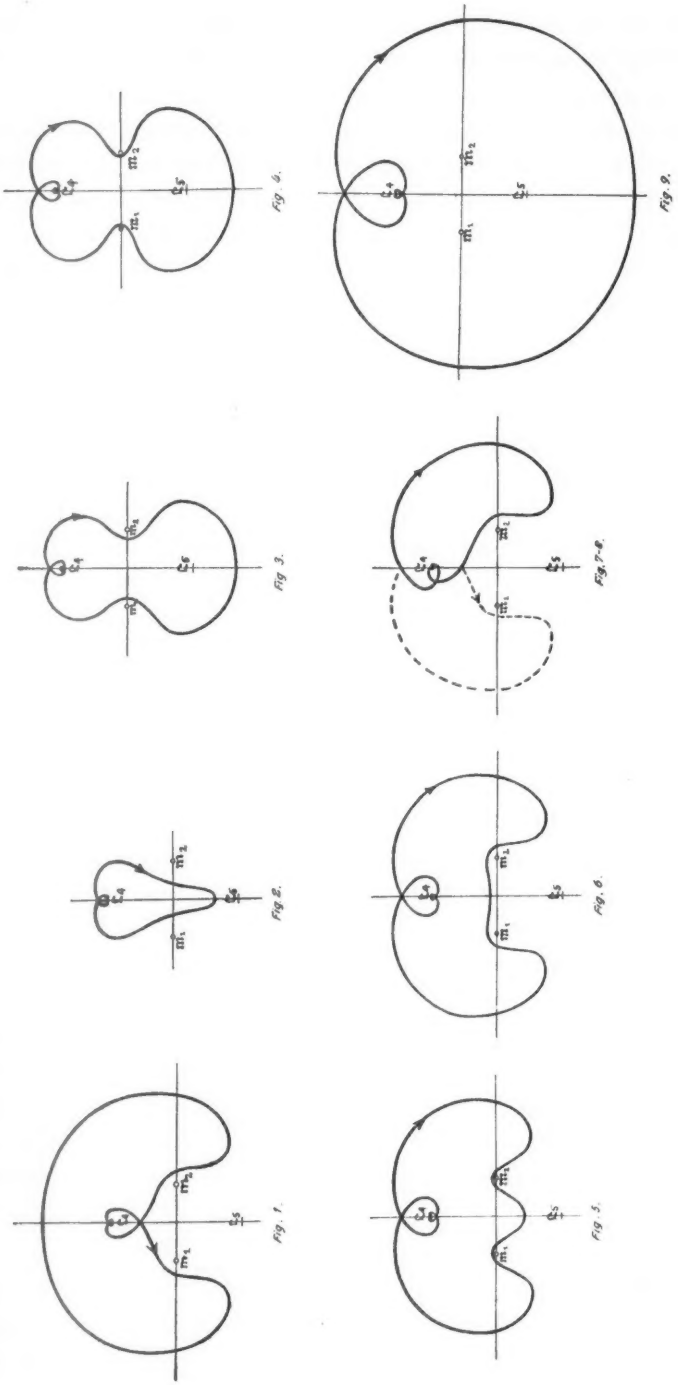


PLATE III.

orbit V in Fig. 10, we find that this orbit is identical with the boundary orbit that was found during the study of the class of orbits mentioned before, which had its origin at infinity.

A more detailed investigation showed that the four other orbits are also boundary orbits of classes of periodic orbits. Thus we have for instance a class of periodic orbits the boundary orbits of which are the orbits III and IV in Fig. 10. Fig. 11 reproduces these two boundary orbits drawn completely. In between these two boundary orbits we have to imagine a class of orbits—infinite in number—the boundary orbits of which are the two asymptotic orbits.

Herewith we have been led to a quite new principle of finding classes of periodic orbits. Formerly the starting point was those periodic orbits, which were known to exist near the masses or the libration points, and through laborious numerical calculations the evolution of each separate class was followed, until the whole evolution of the class up to the natural boundary curves had been covered. Now we are faced with the inverse problem: Once we have found an asymptotic orbit, which is also periodic (with infinitely long period), then we also have a boundary orbit for a whole class of periodic orbits.

We shall not enter into further details, but only touch upon one further point. We found the 5 boundary orbits in Fig. 10 by looking for asymptotic orbits which cut the X-axis at right angles. Of course we can also investigate the problem of finding orbits asymptotic to L_4, which do not cut the X-axis at right angles, but do intersect the Y-axis perpendicularly, one quadrant later. We shall have to deal, in that case, with orbits that are not—like those 5 mentioned before—asymptotic to L_5, but which, having cut the Y-axis, continue in a branch symmetrical to the first with respect to the Y-axis, and end in the libration point L_4, where they started. Our investigations now showed that 6 such orbits exist. They can be found on Plate II, and they are all given together on Plate III (No. 2, 3, 4, 5, 6, and 9). All of these orbits are boundary orbits of classes of periodic orbits, and of course the reasoning can be extended: We can look for all orbits, asymptotic out of L_4, which on cutting the X-axis for the second time are perpendicular to it, and orbits, which on cutting the Y-axis for the second time are at right angles to it, etc., to infinity. In this way we will get boundary orbits of an infinitely increasing number of classes of more and more complicated periodic orbits.

Hill's investigations of the variational curve of the moon and of the problem of the curves of zero velocity introduced new principles into lunar

theory. Poincaré showed the importance of periodic orbits and proved the existence of asymptotic orbits in the problem of three bodies. Thiele-Burrau and G. H. Darwin introduced the methods of numerical integration in the investigations of the restricted problem. In these facts we have the origin of the work on the restricted problem that has been carried out at the Copenhagen Observatory.

ASTRONOMISK OBSERVATORIUM,
KØBENHAVN.

ZUR THEORIE DER DIRICHLETSCHEN REIHEN.

Von OTTO TOEPLITZ.

Die mathematischen Gedanken, die G. W. Hill in seiner berühmten Arbeit über das Mondperigäum entwickelt hat, sind eines der Einfallstore geworden, durch die man den Zugang zum Bereich der unendlich vielen Veränderlichen gefunden hat. Von dieser Arbeit Hills angeregt, hat H. Poincaré die allgemeinen Sätze über unendliche Determinanten entdeckt, die H. von Koch zu einer geschlossenen Theorie abgerundet hat und die I. Fredholm als ein Vorbild bei seiner Konzeption der allgemeinen Theorie der linearen Integralgleichungen dienten. Der Allgemeinheitsgrad dieser Sätze, der in dieser historischen Entwicklung als ein entscheidendes Moment erscheint, ist bei Hill explizite nicht anzutreffen; er handelt scheinbar von einem speziellen Gleichungssystem. Das ist nur scheinbar der Fall. In Wahrheit hat er als ein genialer numerischer Rechner aus den Zahlen seines Spezialfalles die allgemeinen Gedanken herausschauen lassen, die er an der Hand des Sonderfalls erkannte, und vielleicht sind diese Gedanken von der Theorie der unendlich vielen Variablen noch heute nicht voll ausgeschöpft worden.

Obgleich die vorliegende Arbeit zu der oben berührten Aufgabe keinen direkten Bezug hat, soll sie doch Zeugnis ablegen von dem Andenken, das die Lehre von den unendlich vielen Variablen dem Geiste Hills schuldet.

Die Arbeit ist der unveränderte Abdruck von einem Teil eines Manuskripts, das im Jahre 1918 niedergeschrieben worden ist, in den Jahren 1918-1922 einer Reihe von Fachgenossen vorlag und im mathematischen Seminar der Hamburger Universität referiert worden ist. Herr A. Wintner, dem ich für die Anregung zum Abdruck und für die Besorgung der Redaktion besonderen Dank schulde, hat dieses Manuskript im September 1929 kennen gelernt.

1. Die *D*-Formen. Es sei $f(ti)$ eine für $-\infty < t < +\infty$ erklärte Funktion, für welche alle Hadamardschen Koeffizienten[1]

$$c(\lambda) = \mathfrak{M}\{f(ti)e^{i\lambda t}\} \qquad (-\infty < \lambda < +\infty)$$

existieren, wobei $\mathfrak{M}$ die Operation der Mittelbildung bezeichnet:

$$\mathfrak{M}\{\cdots\} = \lim_{\omega=+\infty} \frac{1}{2\omega} \int_{-\omega}^{\omega} \cdots dt.$$

Und zwar soll $f(ti)$ derart sein, dass die Hadamardschen Koeffizienten $c(\lambda)$ ausser für $\lambda = \lg 1, \lg 2, \lg 3, \cdots$ verschwinden. Der Koeffizient $c(\lg n)$

[1] J. Hadamard, *Acta Mathematica*, Bd. 22 (1899), insb. S. 60-63.

werde kurz als c_n bezeichnet, also insbesondere $c(0)$ als c_1. Speziall wird also nach Hadamard und Schnee eine solche Belegung stets dann vorliegen, wenn man eine spezielle Dirichletsche Reihe $\Sigma c_n/n^s$ hat, die auf $s = ti$ absolut oder über $s = ti$ hinaus bedingt konvergiert, und wenn man ihren Summenwert $f(ti)$ als Belegung nimmt.

Alsdann betrachte man den nicht-negativen Ausdruck [2]

$$\begin{aligned}(1)\quad &(\bar{x}_1 + 2^{ti}\bar{x}_2 + 3^{ti}\bar{x}_3 + \cdots + n^{ti}\bar{x}_n)(x_1 + 2^{-ti}x_2 + 3^{-ti}x_3 + \cdots + n^{-ti}x_n)\\ &= (x_1\bar{x}_1 + \cdots + x_n\bar{x}_n) + (x_1\bar{x}_2 + x_2\bar{x}_4 + x_3\bar{x}_6 + \cdots)2^{ti}\\ &\quad + (x_2\bar{x}_3 + x_4\bar{x}_6 + \cdots)(3/2)^{ti} + \cdots + x_1\bar{x}_n n^{ti}\\ &\quad + (\bar{x}_1x_2 + \bar{x}_2x_4 + \bar{x}_3x_6 + \cdots)2^{-ti}\\ &\quad + (\bar{x}_2x_3 + \bar{x}_4x_6 + \cdots)(3/2)^{-ti} + \cdots + \bar{x}_1x_n n^{-ti},\end{aligned}$$

multipliziere ihn mit $f(ti)$ und bilde den Integralmittelwert nach t:

$$(2)\quad D_n(x,\bar{x}) = \mathfrak{M}\{f(ti)(\bar{x}_1 + 2^{ti}\bar{x}_2 + \cdots + n^{ti}\bar{x}_n)(x_1 + 2^{-ti}x_2 + \cdots + n^{-ti}x_n)\}.$$

Nun ist vorausgesetzt, dass

$$\mathfrak{M}(f(ti)n^{ti}) = c_n \text{ oder } = 0,$$

je nachdem n positiv-ganz ist oder nicht. Führt man also in (2) unter dem Integralzeichen statt der linken Seite von (1) die rechte Seite ein, so verschwindet die Mehrzahl der Summanden, und es bleibt:

$$\begin{aligned}(3)\quad D_n(x,\bar{x}) &= c_1(x_1\bar{x}_1 + \cdots + x_n\bar{x}_n) + c_2(x_1\bar{x}_2 + x_2\bar{x}_4 + x_3\bar{x}_6 + \cdots)\\ &\quad + c_3(x_1\bar{x}_3 + x_2\bar{x}_6 + \cdots) + \cdots + c_n(x_1\bar{x}_n).\end{aligned}$$

Ist nun überdies die Belegung $f(ti)$ beschränkt, $|f(ti)| \leqq M$, so kann man auf das Integral in (2) den ersten Mittelwertsatz der Integralgleichung anwenden, da der Faktor von $f(ti)$ eine reelle, nicht-negative Funktion ist, und erhält

$$|D_n(x,\bar{x})| \leqq M\mathfrak{M}\{(\bar{x}_1 + 2^{ti}\bar{x}_2 + \cdots + n^{ti}\bar{x}_n)(x_1 + 2^{-ti}x_2 + \cdots + n^{-ti}x_n)\},$$

also unter Benutzung der Relation (1):

$$|D_x(x,\bar{x})| \leqq M(x_1\bar{x}_1 + \cdots + x_n\bar{x}_x),$$

da $\mathfrak{M}\{e^{i\lambda t}\}$ verschwindet oder gleich eins wird, je nachdem $\lambda \neq 0$ oder $\lambda = 0$ ist.

THEOREM I. *Befriedigt eine Belegung ausser den eingangs dieses Paragraphen aufgeführten Voraussetzungen noch die weitere* $|f(ti)| \leqq M$, *und ist* H_n *das Maximum des Ausdrucks* (3) *unter der Nebenbedingung* $x_1\bar{x}_1 + \cdots + x_n\bar{x}_n = 1$, *so ist* $H_n \leqq M$ *für jedes* n.

[2] Hierbei sind die Summen in den Klammern alle soweit zu erstrecken, als die Indizes $\leqq n$ bleiben.

BEMERKUNG. Bedient man sich der Sprechweise der Theorie der unendlich vielen Variablen, so ist $D_n(x, \bar{x})$ nichts anderes als der n-te Abschnitt der Bilinearform von unendlich vielen Variablen

$$\mathrm{D}(x, y) = c_1(x_1y_1 + x_2y_2 + \cdots) + c_2(x_1y_2 + x_2y_4 + \cdots) + c_3(x_1y_3 + x_2y_6 + \cdots) + \cdots,$$

nur dass in den n-ten Abschnitt $D_n(x, y)$ hiervon noch $y_1 = \bar{x}_1, \cdots, y_n = \bar{x}_n$ gesetzt ist. Nun besagt ein einfacher Hilfssatz[3] über Bilinearformen von endlich vielen Veränderlichen $A(x, y)$, dass das Maximum von $|A(x, y)|$ für $\Sigma x_a\bar{x}_a = 1$, $\Sigma y_a\bar{y}_a = 1$ höchstens gleich dem Doppelten des Maximums von $A(x, \bar{x})$ für $\Sigma x_a\bar{x}_a = 1$ ist; also liegt im vorliegenden Falle $|D_n(x, y)|$ unter $2M$, und Theorem I kann auch so formuliert werden:

Befriedigt eine Belegung die eingangs dieses Paragraphen aufgeführten Voraussetzungen, so soll die aus ihren Hadamardschen Koeffizienten c_n aufgebaute Bilinearform $D(x, y)$ als die zu der Belegung gehörige *„D-Form"* bezeichnet werden. *Ist die Belegung beschränkt, $|f(ti)| \leqq M$, so ist die D-Form im Hilbertschen Sinne beschränkt, und ihre obere Schranke liegt unter $2M$.*

Die Koeffizientenmatrix einer solchen D-Form hat die Eigentümlichkeit, dass die Stellen, die den nämlichen Koeffizienten tragen, je eine gerade Linie erfüllen und dass alle diese geraden Linien ein Strahlenbüschel bilden mit endlichem Zentrum:[4]

$$D = \left\| \begin{array}{cccccccccccccc}
c_1 & c_2 & c_3 & c_4 & c_5 & c_6 & c_7 & c_8 & c_9 & c_{10} & c_{11} & c_{12} & \cdot & \cdot & \cdot \\
0 & c_1 & 0 & c_2 & 0 & c_3 & 0 & c_4 & 0 & c_5 & 0 & c_6 & \cdot & \cdot & \cdot \\
0 & 0 & c_1 & 0 & 0 & c_2 & 0 & 0 & c_3 & 0 & 0 & c_4 & \cdot & \cdot & \cdot \\
0 & 0 & 0 & c_1 & 0 & 0 & 0 & c_2 & 0 & 0 & 0 & c_3 & \cdot & \cdot & \cdot \\
0 & 0 & 0 & 0 & c_1 & 0 & 0 & 0 & 0 & c_2 & 0 & 0 & \cdot & \cdot & \cdot \\
0 & 0 & 0 & 0 & 0 & c_1 & 0 & 0 & 0 & 0 & 0 & c_2 & \cdot & \cdot & \cdot \\
0 & 0 & 0 & 0 & 0 & 0 & c_1 & 0 & 0 & 0 & 0 & 0 & \cdot & \cdot & \cdot \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & c_1 & 0 & 0 & 0 & 0 & \cdot & \cdot & \cdot \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & c_1 & 0 & 0 & 0 & \cdot & \cdot & \cdot \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & c_1 & 0 & 0 & \cdot & \cdot & \cdot \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & c_1 & 0 & \cdot & \cdot & \cdot \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & c_1 & \cdot & \cdot & \cdot \\
\cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\
\cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\
\cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot
\end{array} \right\|.$$

[3] Vgl. eine inzwischen in der *Mathematischen Zeitschrift* (Bd. 2 (1918), S. 187) erschienene Arbeit.

[4] Während sie bei den L-Formen Parallelstrahlenbüschel bilden.

Das System dieser D-Formen oder D-Matrizen ist isomorph den rein formal betrachteten zugehörigen Dirichletschen Reihen, d. h. der Summe und dem formalen Produkt zweier spezieller Dirichletscher Reihen gehört die Summe bzw. das Produkt der zugehörigen D-Formen (im Sinne des Matrizenkalküls) zu.

Nach einem bekannten Hilfssatze von Weierstrass kann man aus der Positivität des Faktors von $f(ti)$ in (2) mehr entnehmen:

ZUSATZ. *Ist $\mathfrak{R}$ der kleinste konvexe Bereich, der alle Werte der Belegung $f(ti)$ umschliesst, so liegt jeder Wert, den $D_n(x, \bar{x})$ für $\Sigma x_\alpha \bar{x}_\alpha = 1$ annimmt, in $\mathfrak{R}$.*

2. Umkehrung von Theorem I für absolut konvergente Dirichletsche Reihen. Ist $f(s) = \Sigma c_n/n^s$ absolut konvergent für $s = ti$, ist m. a. W. $\Sigma \mid c_n \mid$ konvergent, so ist $\Sigma c_n/n^s$ a fortiori gleichmässig konvergent, und es folgt aus den Ergebnissen von Hadamard, dass $f(ti)$ alle Voraussetzungen von § 1, auch die der Beschränktheit, erfüllt. Es gilt dann:

THEOREM II. *Ist $\Sigma \mid c_n \mid$ konvergent und wird $\Sigma c_n/n^{ti}$ gesetzt, so kommt der in Theorem I mit $D_n(x, \bar{x})$ bezeichnete Ausdruck unter der Nebenbedingung $\Sigma x_\alpha \bar{x}_\alpha = 1$ und bei hinreichend grossem n jedem Wert der Belegung $f(ti)$ beliebig nahe.*

Der Beweis werde mit dem naheliegenden Versuch begonnen

$$x_1 = 1/\sqrt{\alpha}, \cdots, x_\alpha = 1/\sqrt{\alpha}, \text{ alle weiteren } x_n = 0. \tag{1}$$

Dann ist, wenn $n \geqq \alpha$, $x_1\bar{x}_1 + \cdots + x_n\bar{x}_n = 1$. Ferner kommt in $D_n(x, \bar{x})$ offenbar c_1 genau α mal vor, jedesmal mit dem Faktor $1/\sqrt{\alpha} \cdot 1/\sqrt{\alpha} = 1/\alpha$, das gibt zusammen c_1, während c_2 genau $\frac{1}{2}\alpha$ mal vorkommt, wenn α gerade ist, und $\frac{1}{2}(\alpha - 1)$ mal bei ungeradem α, u. s. f.; also ist für das Wertsystem (1)

$$D_n(x, \bar{x}) = c_1 + \tfrac{1}{2}c_2 + \tfrac{1}{3}c_3 + \cdots + \frac{1}{\alpha}c_\alpha + \Delta_\alpha,$$

wo die kleine Abweichung Δ_α jedenfalls der Abschätzung unterliegt

$$\mid \Delta_\alpha \mid \leqq \mid c_2 \mid \frac{1}{2\alpha} + \mid c_3 \mid \tfrac{2}{3} \cdot \frac{1}{\alpha} + \cdots + \mid c_\alpha \mid \frac{1}{\alpha} \leqq \frac{\mid c_2 \mid + \cdots + \mid c_\alpha \mid}{\alpha}.$$

Wegen der Konvergenz von $\Sigma \mid c_\alpha \mid$ ist $\lim\limits_{\alpha=\infty} \mid c_\alpha \mid = 0$, also nach dem Satz vom arithmetischen Mittel auch $\lim\limits_{\alpha=\infty} \Delta_\alpha = 0$, und $D_n(x, \bar{x})$ nähert sich für das Wertsystem (1) und mit wachsendem α und $n \geqq \alpha$ immer mehr dem Werte

$$c_1 + \tfrac{1}{2}c_2 + \tfrac{1}{3}c_3 + \cdots = f(1). \tag{2}$$

So könnte man von $f(1 + ti)$, wie eben für $f(1)$, dasjenige einsehen, was von $f(0)$ und $f(0 + ti)$ gezeigt werden soll.

Es soll deshalb der Versuch mit anderen, komplizierteren Wertsystemen x_n

wiederholt werden, die übrigens den Vorteil haben werden, dass die lästige Abschätzung eines Δ_a nicht auftritt. Es sei zuerst, mit $|\eta| = 1$,

$$(3) \qquad x_1 = 1/\sqrt{\alpha},\ x_2 = \eta/\sqrt{\alpha},\ x_4 = \eta^2/\sqrt{\alpha}, \cdots, x_{2^{a-1}} = \eta^{a-1}/\sqrt{\alpha},$$

alle anderen $x_\nu = 0$.

So erhält man neben $\sum_{\nu=1}^{n} x_\nu \bar{x}_\nu = 1$, wenn n hinreichend gross ist, $n \geqq 2^a$,

$$(4) \qquad D_n(x, \bar{x}) = c_1 + \frac{\alpha-1}{\alpha}\bar{\eta}c_2 + \frac{\alpha-2}{\alpha}\bar{\eta}^2 c_4 + \cdots + \frac{1}{\alpha}\bar{\eta}^{a-1}c_{2^{a-1}},$$

oder, wenn $\eta = 2^{ti}$ gesetzt wird,

$$(5) \qquad D_n(x, \bar{x}) = c_1 + \frac{\alpha-1}{\alpha}\frac{c_2}{2^{ti}} + \frac{\alpha-2}{\alpha}\frac{c_4}{4^{ti}} + \cdots + \frac{1}{\alpha}c_{2^{a-1}}\frac{1}{2^{(a-1)ti}},$$

d. h. das α-te arithmetische Mittel der unendlichen Reihe

$$(6) \qquad c_1 + \frac{c_2}{2^{ti}} + \frac{c_4}{4^{ti}} + \frac{c_8}{8^{ti}} + \cdots.$$

Da diese eine Teilreihe der als absolut konvergent vorausgesetzten Reihe $\Sigma c_n/n^{ti}$ ist, ist sie auch konvergent, und somit auch die Folge ihrer arithmetischen Mittel, und diese konvergiert gegen denselben Wert. Dem Wert dieser Teilreihe kommt aber zunächst die Form $D_n(x, \bar{x})$ im angegebenen Sinne beliebig nahe.

Es sei nun weiter bei hinreichend hohem n $(n \geqq 2^a \cdot 3^\beta)$:

$$|\eta| = 1, \qquad |\theta| = 1$$

und

$$(7) \quad \begin{array}{llllll} x_1 = \dfrac{1}{\sqrt{\alpha\beta}}, & x_2 = \dfrac{\eta}{\sqrt{\alpha\beta}}, & x_4 = \dfrac{\eta^2}{\sqrt{\alpha\beta}}, \cdots, & x_{2^{a-1}} = \dfrac{\eta^{a-1}}{\sqrt{\alpha\beta}}, \\ x_3 = \dfrac{\theta}{\sqrt{\alpha\beta}}, & x_{3.2} = \dfrac{\theta\eta}{\sqrt{\alpha\beta}}, & x_{3.4} = \dfrac{\theta\eta^2}{\sqrt{\alpha\beta}}, \cdots, & x_{3.2^{a-1}} = \dfrac{\theta\eta^{a-1}}{\sqrt{\alpha\beta}}, \\ \cdots\cdots & \cdots\cdots & \cdots\cdots & \cdots\cdots \\ x_{3^{\beta-1}} = \dfrac{\theta^{\beta-1}}{\sqrt{\alpha\beta}}, & x_{3^{\beta-1}.2} = \dfrac{\theta^{\beta-1}\eta}{\sqrt{\alpha\beta}}, & x_{3^{\beta-1}.4} = \dfrac{\theta^{\beta-1}\eta^2}{\sqrt{\alpha\beta}}, \cdots, & x_{3^{\beta-1}.2^{a-1}} = \dfrac{\theta^{\beta-1}\eta^{a-1}}{\sqrt{\alpha\beta}}, \end{array}$$

alle anderen $x_\nu = 0$, so ist wiederum $\Sigma x_\nu \bar{x}_\nu = 1$, und $D_n(x, \bar{x})$ wird gleich

$$(8) \quad \begin{array}{l} c_1 + c_2\bar{\eta}\dfrac{(\alpha-1)\beta}{\alpha\beta} + c_4\bar{\eta}^2\dfrac{(\alpha-2)\beta}{\alpha\beta} + \cdots + c_{2^{a-1}}\bar{\eta}^{a-1}\dfrac{\beta}{\alpha\beta} \\ + c_3\bar{\theta}\dfrac{\alpha(\beta-1)}{\alpha\beta} + c_{3.2}\bar{\theta}\bar{\eta}\dfrac{(\alpha-1)(\beta-1)}{\alpha\beta} + c_{3.4}\bar{\theta}\bar{\eta}^2\dfrac{(\alpha-2)(\beta-1)}{\alpha\beta} \\ + \cdots + c_{3.2^{a-1}}\bar{\theta}\bar{\eta}^{a-1}\dfrac{\beta-1}{\alpha\beta} \\ + \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ + c_{3^{\beta-1}}\bar{\theta}^{\beta-1}\dfrac{\alpha}{\alpha\beta} + c_{3^{\beta-1}.2}\bar{\theta}^{\beta-1}\bar{\eta}\dfrac{\alpha-1}{\alpha\beta} + c_{3^{\beta-1}.4}\bar{\theta}^{\beta-1}\bar{\eta}^2\dfrac{\alpha-2}{\alpha\beta} \\ + \cdots + c_{3^{\beta-1}.2^{a-1}}\bar{\theta}^{\beta-1}\bar{\eta}^{a-1}\dfrac{1}{\alpha\beta}. \end{array}$$

Nun ist die Doppelreihe

$$
\begin{aligned}
&c_1 + c_2/2^{ti} + c_4/4^{ti} + \cdots \\
&+ c_3 + c_6/6^{ti} + c_{12}/12^{ti} + \cdots \\
&+ c_9 + c_{18}/18^{ti} + c_{36}/36^{ti} + \cdots \\
&+ \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot
\end{aligned} \tag{9}
$$

absolut konvergent als Teilreihe der absolut konvergenten Reihe $\Sigma c_n/n^{ti}$. Setzt man also $\theta = 3^{ti}$, $\eta = 2^{ti}$, so kann man zuerst bei festem β den Index α so gross wählen, dass (8) beliebig wenig abweicht von

$$
\begin{aligned}
&\{c_1 + c_2(1/2^{ti}) + c_4(1/4^{ti}) + \cdots\} \\
&+[(\beta-1)/\beta]\{c_3(1/3^{ti}) + c_{3.2}(1/6^{ti}) + c_{3.4}(1/12^{ti}) + \cdots\} \\
&+ \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \\
&+(1/\beta)\{c_{3^{\beta-1}}[1/(3^{\beta-1})^{ti}] + c_{3^{\beta-1}.2}[1/(3^{\beta-1}.2)^{ti}] + c_{3^{\beta-1}.4}[1/(3^{\beta-1}.4)^{ti}] + \cdots\},
\end{aligned}
$$

d. h. von dem β-ten arithmetischen Mittel aus den Zeilensummen der Doppelreihe (9). Endlich kann man also β so gross werden lassen, dass man den Summenwert der gesamten Doppelreihe (9) beliebig nahe kommt. Der Wert von (9) wird somit durch Werte von $D_n(x, \bar{x})$ bei hinreichend hohem n approximiert; (9) ist aber schon eine umfassendere Teilreihe von $\Sigma c_n/n^{ti}$ als (6).

Setzt man allgemeiner, indem man mit p_ν die ν-te Primzahl bezeichnet, $|\eta_k| = 1$ für $k = 1, \cdots, \nu$,

$$x_{p_1^{u_1} \ldots p_\nu^{u_\nu}} = \eta_1^{u_1} \cdots \eta_\nu^{u_\nu}/\sqrt{\alpha_1 \cdots \alpha_\nu} \text{ für } 0 \leqq u_1 < \alpha_1, \cdots, 0 \leqq u_\nu < \alpha_\nu,$$

alle anderen $x_k = 0$, so folgt durch eine analoge Schlussweise, dass man mit $D_n(x, \bar{x})$ bei hinreichend hohem n diejenige Teilreihe von $\sum_{n=1}^{\infty} c_n/n^{ti}$ approximiert, deren Glieder einen nur durch $p_1, \cdots, p_\nu$ teilbaren Index n haben. Nun ist die ganze Reihe absolut konvergent; man kann also N so gross wählen, dass $\sum_{n=N+1}^{\infty} |c_n/n^{ti}| \leqq \frac{1}{2}\epsilon$ ist. Alsdann wähle man ν so gross, dass alle in den Zahlen $1, 2, 3, \cdots, N$ aufgehenden Primfaktoren unter $p_1, \cdots, p_\nu$ enthalten sind. Dann erhält man als Wert von $D(x, \bar{x})$ eine Teilreihe, die zunächst *alle* Glieder von $\sum_{n=1}^{N} c_n/n^{ti}$ enthält und dann noch manche von dem Rest; die letzteren zusammen aber betragen weniger als $\frac{1}{2}\epsilon$; also ist $\sum_{n=1}^{\infty} c_n/n^{ti}$ von einem Wert von $D_n(x, \bar{x})$ um weniger als ϵ entfernt, q. e. d.

Die Grössen $H_1, H_2, \cdots$ bilden eine Folge zunehmender, positiver Zahlen; nach Theorem I liegen sie alle unter M; also haben sie einen Grenzwert

$\lim_{n=\infty} \mathrm{H}_n = \mathrm{H} \leqq M$. Theorem II gestattet nun, wenn M die genaue obere Grenze von $|f(ti)|$ ist, zu folgern, dass $\mathrm{H} = M$ ist:

FOLGERUNG. *Ist M die obere Grenze des Betrages von $\Sigma c_n/n^{ti} = f(ti)$ und $\lim_{n=\infty} \mathrm{H}_n = \mathrm{H}$, so ist $\mathrm{H} = M$.*

3. Das Vertikalenmaximum. Die bisher aufgestellten Theoreme gestatten bereits einige Anwendungen auf die Theorie der Dirichletschen Reihen.

DEFINITION. Ich sage, eine Belegung $\phi(t)$, $-\infty < t < +\infty$, nimmt einen Wert „asymptotisch" an, wenn sie ihm ausserhalb jedes Intervalls $-T \leqq t \leqq +T$ beliebig nahe kommt. Im gleichen Sinne spreche ich von dem „asymptotischen Maximum" von $|\phi(t)|$, von dem „asymptotischen Wertvorrat" u. s. w.

Auf Grund dieser Terminologie kann man eine Verschärfung, deren Theorem I fähig ist, folgendermassen formulieren:

THEOREM III. *Ist $f(ti)$ eine Belegung, die den Voraussetzungen am Eingang von § 1 genügt, ist $\tilde{M}$ das asymptotische Maximum von $|f(ti)|$ und $\mathrm{H} = \lim \mathrm{H}_n$ dieselbe Grösse wie in Theorem* I, *so ist $\mathrm{H} \leqq \tilde{M} \leqq M$.*

ZUSATZ. *Ist* K *der kleinste konvexe Bereich, der den asymptotischen Wertvorrat von $f(ti)$ enthält, so liegen alle Werte, die $D_n(x, \bar{x})$ für $\Sigma x_\alpha \bar{x}_\alpha = 1$ annimmt, in* K (*das ganz in* $\mathfrak{R}$ *enthalten ist*).

Beweis. Ist $\tilde{M}$ das asymptotische Maximum von $|f(ti)|$, so besagt dies: wenn $\epsilon > 0$ beliebig vorgegeben ist, kann man ein solches Intervall $-T \leqq t \leqq +T$ finden, dass die obere Grenze von $|f(ti)|$ ausserhalb dieses Intervalls $\leqq \tilde{M} + \epsilon$ ist. Betrachtet man nun neben der Belegung $f(ti)$ die andere $g(ti)$, die für $-T \leqq t \leqq +T$ den Wert 0 hat, ausserhalb dieses Intervalls aber mit $f(ti)$ übereinstimmt, so ist die obere Grenze von $|g(ti)| \leqq M + \epsilon$. Dann existieren aber die Hadamardschen Koeffizienten von $g(ti)$ und sind genau dieselben wie die von $f(ti)$; es wird also auch zu $g(ti)$ dieselbe D-Form gehören, wie zu $f(ti)$, mit derselben oberen Grenze H. Wendet man also Theorem I auf $g(ti)$ an, so erhält man $\mathrm{H} \leqq \tilde{M} + \epsilon$, und da ϵ beliebig klein gewählt werden konnte, $\mathrm{H} \leqq \tilde{M}$.

Nun ist unter den Voraussetzungen von Theorem II $\mathrm{H} = M$; nach Theorem III ist $\mathrm{H} \leqq \tilde{M}$, und seiner Definition nach ist $\tilde{M} \leqq M$; also ist auch $\mathrm{H} = \tilde{M}$:

SATZ 1. *Im Gebiete absoluter Konvergenz ist für eine Dirichletsche Reihe $f(s) = \Sigma c_n/n^s$ längs jeder Vertikalen $s = \sigma + ti$ die obere Grenze von*

$|f(\sigma+ti)|$ *mit dem asymptotischen Maximum* $\tilde{M}(\sigma)$ *dieser Belegung identisch. Genauer liegen alle Werte von* $f(\sigma+ti)$ *innerhalb des kleinsten konvexen Bereiches* $\mathrm{K}(\sigma)$, *den man um den asymptotischen Wertvorrat von* $f(\sigma+ti)$ *legen kann.*

Dieser Satz enthält als Folgerungen einige Theoreme, die Herr Bohr 1910 bewiesen hat. Für $\zeta(s)=\Sigma 1/n^s$ nämlich ist $\Re(s)>1$ das Gebiet absoluter Konvergenz, und da die Koeffizienten reell-positiv sind, ist $|\zeta(\sigma+ti)|\leqq\zeta(\sigma)$, also $\tilde{M}(\sigma)=\zeta(\sigma)$. Satz 1 zeigt nun, dass $\zeta(\sigma+ti)$ für beliebig hohe t dem Wert $\zeta(\sigma)$ beliebig nahe kommt. Da nun $\zeta(\sigma)$ unbegrenzt wächst, wenn σ von rechts an 1 heranrückt, ist also $\zeta(s)$ in der Halbebene $\Re(s)>1$ unbeschränkt. Entsprechendes gilt für jede Reihe $\Sigma c_n/n^s$ mit lauter nichtnegativen reellen Koeffizienten.

SATZ 2. *Innerhalb des Gebietes absoluter Konvergenz ist das Vertikalenmaximum* $M(\sigma)$ *eine nirgends ansteigende Funktion der Vertikalenabscisse* σ.

BEWEIS. Das Vertikalenmaximum $M(\sigma)$ ist nach dem Ende von § 2 identisch mit der oberen Grenze $\mathrm{H}(\sigma)$ der D-Form mit den Koeffizienten c_n/n^σ:

$$D(x,\bar{x};\sigma)=c_1(x_1\bar{x}_1+x_2\bar{x}_2+\cdots)+\frac{c_2}{2^\sigma}(x_1\bar{x}_2+x_2\bar{x}_4+\cdots)$$
$$+\frac{c_3}{3^\sigma}(x_1\bar{x}_3+x_2\bar{x}_6+\cdots)+\cdots$$

für $\Sigma x_a\bar{x}_a=1$. Nun geht diese Form durch die Substitution $x_n=n^{-\epsilon}\xi_n$ in $D(\xi,\bar{\xi};\sigma+2\epsilon)$ über, d. h. in die D-Form zu der um 2ϵ weiter nach rechts gelegenen Vertikalen. Es ist $\Sigma x_n\bar{x}_n=\Sigma n^{-2\epsilon}\xi_n\bar{\xi}_n$. Ist also $\Sigma\xi_n\bar{\xi}_n\leqq 1$, so ist erst recht $\Sigma x_n\bar{x}_n\leqq 1$, aber nicht umgekehrt. Ist also $|D(x,\bar{x};\sigma)|$ für $\Sigma x_n\bar{x}_n\leqq 1$ unter $M(\sigma)$ gelegen, so wird a fortiori $|D(\xi,\bar{\xi},\sigma+2\epsilon)|$ für $\Sigma\xi_n\bar{\xi}_n\leqq 1$ unter $M(\sigma)$ liegen, d. h. es ist $M(\sigma+2\epsilon)\leqq M(\sigma)$.

4. Umkehrung von Theorem I ohne einschränkende Annahme. Es sei eine D-Form $D(x,y)$ rein formal vorgelegt, und es sei lediglich von ihr vorausgesetzt, dass sie beschränkt ist. Daraus folgt an sich nach den allerersten Sätzen der Theorie der unendlich vielen Veränderlichen, dass die Quadratsumme $\Sigma|c_n|^2$ der Beträge ihrer Koeffizienten konvergiert; es sei aber keinerlei weitere Annahme, wie z. B. die der Konvergenz von $\Sigma|c_n|$ hinzugefügt. Alsdann kann man sich, anknüpfend an die Betrachtungen von § 2, derselben Methode bedienen, die Herr Bohr,[5] auf anderen Voraussetzungen fussend, zu einem ähnlichen Ziel benutzt hat.

Zuerst folgt aus der im § 1 im Anschluss an Theorem I gemachten Bemerkung, dass alle $|c_n| \leqq 2H$ sind, wo H die obere Grenze von $|D(x, \bar{x})|$ für $\Sigma x_a \bar{x}_a = 1$ ist. Sei nun, mit Herrn Bohr, zunächst rein formal:

$$f_1(s) = \sum_{n=1}^{\infty} c_{2^n} \frac{1}{2^{ns}}, \qquad f_2(s) = \sum_{n_1=1}^{\infty} \sum_{n_2=1}^{\infty} c_{2^{n_1}3^{n_2}} \frac{1}{2^{n_1 s} \cdot 3^{n_2 s}}, \cdots,$$

$$f_\nu(s) = \sum_{n_1=1}^{\infty} \cdots \sum_{n_\nu=1}^{\infty} c_{p_1^{n_1} \ldots p_\nu^{n_\nu}} \frac{1}{p_1^{n_1 s} \cdots p_\nu^{n_\nu s}}, \cdots,$$

so konvergiert also $f_1(s)$ für $\Re(s) > 0$ absolut und stellt daselbst eine analytische Funktion dar, ebenso $f_2(s), f_3(s), \cdots$.

Ferner folgt aus der schon erwähnten Konvergenz von $\Sigma |c_n|^2$ mit Hilfe der Schwarzschen Ungleichung, dass $\Sigma |c_n|/n^{\frac{1}{2}+\epsilon}$ für jedes $\epsilon > 0$ konvergiert, dass also $\Sigma c_n/n^s$ für $\Re(s) > \frac{1}{2}$ absolut konvergiert.

Endlich besagen (5) und (6) von § 2, dass das α-te arithmetische Mittel der Reihe $f_1(s)$ für $s = ti$ unter H gelegen ist; also ist $|f_1(s)| \leqq H$ für $\Re(s) > 0$, und ebenso folgt aus § 2, dass $|f_\nu(s)| \leqq H$ für $\Re(s) > 0$ und jedes $\nu = 1, 2, \cdots$.

Ist nun $1 + \tau i$ irgendein Punkt auf der Vertikalen mit der Abszisse 1, so betrachte man einen Kreis vom Radius $\frac{1}{2}$ um diesen Punkt. Innerhalb dieses Kreises ist $\Sigma c_n/n^s$ absolut konvergent; $f_1(s), f_2(s), \cdots$ sind Teilreihen davon; dieselben konvergieren also gleichmässig gegen $f(s)$. Nun besagt ein von Stieltjes stammender funktionentheoretischer Hilfssatz: wenn eine Folge analytischer Funktionen $f_1(s), f_2(s), \cdots$ in einem Kreise gleichmässig gegen $f(s)$ konvergiert und wenn dieselben Funktionen $f_\nu(s)$ in einem grösseren Gebiet dem Betrage nach unter einer gemeinsamen festen Grenze liegen, so konvergieren sie in jedem beschränkten abgeschossenen Teilbereich dieses grösseren Gebietes gleichmässig. Im vorliegenden Falle ergibt sich daraus, dass die obigen Funktionen $f_\nu(s)$ auch im Kreise vom Radius 1 um den Punkt $1 + \tau i$ gleichmässig konvergieren; die Grenzfunktion wird daselbst die analytische Fortsetzung von $f(s)$ sein und dem Betrage nach unter H liegen.

THEOREM IV. *Ist die zur Reihe $\Sigma c_n/n^s$ gehörige D-Form im Hilbertschen Sinne beschränkt und* H *die obere Grenze von $D(x, \bar{x})$, so konvergiert die Reihe selbst absolut für $\Re(s) > \frac{1}{2}$ und stellt eine analytische Funktion $f(s)$ dar, die bis an $\Re(s) = 0$ heran fortsetzbar ist und dem Betrage nach überall unter* H *liegt.*

BONN, GERMANY.

[5] H. Bohr, *Göttinger Nachrichten*, 1913.

MEAN MOTION.

By HERMANN WEYL.

The problem of mean motion which arose from the theory of the secular perturbations of the planetary orbits [1] and which will here find its complete solution, deals with the *superposition of a finite number of epicycles* or of simple oscillations:

$$z = \sum_{k=1}^{n} a_k e(\vartheta_k), \tag{1}$$

$$\vartheta_k = \vartheta_k(t) = \lambda_k t + \alpha_k. \tag{2}$$

We use the abbreviation $e^{2\pi i\vartheta} = e(\vartheta)$. The *amplitudes* $a_k > 0$, *frequencies* λ_k, and *initial phases* α_k, are real constants; the *time* t is a real variable ranging from $-\infty$ to $+\infty$. The n-dimensional space of the real phases ϑ_k is to be construed as a *torus* Θ by identifying two points $(\vartheta) = (\vartheta_1, \cdots, \vartheta_n)$ and $(\vartheta') = (\vartheta'_1, \cdots, \vartheta'_n)$ which are congruent modulo 1:

$$\vartheta_k \equiv \vartheta'_k \pmod 1.$$

z is first a function of the n phases ϑ_k and then, after the substitution (2), a function of t. With the customary notation, polar coördinates are used in the z-plane:

$$z = r \cdot e(\phi).$$

Let us begin by discussing the phase function. (1) may be interpreted as depicting a (plane open) *linkage* $A_0A_1 \cdots A_n$ consisting of the links

$$A_{k-1}A_k = a_k e(\vartheta_k) \tag{3}$$

of fixed lengths a_k. $z = A_0A_n$ is the vector *spanned* by the linkage. Those points on the torus Θ for which $z = 0$ form what we call the *singular manifold*; it consists of all possible states of a *closed* n-linkage with the prescribed sides a_k. E being any area in the z-plane, the volume $v(E)$ of that part of the torus where $z(\vartheta_1 \cdots \vartheta_n)$ takes on values in E, indicates the phase probability of z lying in E. On account of central symmetry, it suffices to determine the probability

$$W(r) = W(r; a_1 \cdots a_n)$$

of the span z being in absolute value $< r$. By its very nature $W(r)$ is an

increasing function of r varying between the limits 0 and 1. The strip $0 \leqq W \leqq 1$ in a (W, ϕ)-plane is a cylinder when points (W, ϕ), (W, ϕ') are regarded equal if $\phi \equiv \phi' \pmod 1$. The phase probability $v(E)$ is the area of the image of E on the (W, ϕ)-cylinder produced by the mapping

$$z = r \cdot e(\phi) \rightarrow (W(r), \phi).$$

A straightforward analysis by Fourier transforms, carried out by A. Wintner [2], results, for $n \geqq 2$, in the absolutely convergent expression

$$W(r; a_1 \cdot \cdot \cdot a_n) = r \int_0^\infty J_1(r\rho) \prod_{k=1}^{n} J_0(a_k \rho) d\rho \tag{4}$$

in terms of the Bessel functions J_0, J_1. The relationship to the random walk problem is obvious [3].

We now turn to the second standpoint with t as the independent variable, and assume that the λ_k satisfy no homogeneous linear relation

$$\sum_k m_k \lambda_k = 0 \tag{5}$$

with integral coefficients m_k. The *law of equidistribution* (Kronecker-Weyl) [4] then states that the straight line (2) fills the torus with uniform density. In other words, for any piece G of the torus with a Jordan volume G, the relative frequency with which the point $\vartheta(t)$, moving uniformly along a straight line, visits G, tends to G if the period of observation becomes infinitely large, or: *time-probability = phase-probability.* A slightly more general form of the proposition asserts that for any bounded Riemann integrable function $f(\vartheta_1 \cdot \cdot \cdot \vartheta_n)$

$$\frac{1}{t_2 - t_1} \int_{t_1}^{t_2} f((\vartheta(t)))\, dt \text{ tends to } \int_0^1 \cdot \cdot \cdot \int_0^1 f((\vartheta))\, d\vartheta_1 \cdot \cdot \cdot d\vartheta_n \tag{6}$$

with $t_2 - t_1 \rightarrow \infty$, or that the *time-average* of f *equals its phase-average.* Hence the considerations of the preceding paragraph determine the relative frequencies with which the "planet" $z(t)$ visits the various parts E of the z-plane.[1]

However, the problem of mean motion is slightly more intricate. To avoid ambiguities, let us suppose that the planet never passes through the origin (or that the straight line $\vartheta(t)$ does not hit the singular manifold). Then the continuous increment $\phi(t_2) - \phi(t_1)$ of the azimuth $\phi(t)$ of $z(t)$

[1] This is to be interpreted in a figurative sense, as in the astronomical application based on the Copernican rather than the Ptolomaic system $\phi(t)$ is not the longitude of the planet itself, but of its perihelion or its ascending node.

over any time interval $t_1 \leqq t \leqq t_2$ is uniquely determined. ϕ has a *mean motion* μ if

$$\frac{\phi(t_2) - \phi(t_1)}{t_2 - t_1} \to \mu \text{ for } t_2 - t_1 \to \infty.$$

We shall see that such a mean motion μ exists, depending linearly on the frequencies λ_k:

$$\mu = W_1\lambda_1 + \cdots + W_n\lambda_n. \tag{7}$$

The complete statement is as follows.

THEOREM. *Assuming the frequencies λ_k to satisfy no linear relation* (5) *with integral coefficients m_k of sum* 0, *the azimuth ϕ of the superposition* (1), (2) *of epicycles has a mean motion*

$$\mu = W_1\lambda_1 + \cdots + W_n\lambda_n. \tag{7}$$

The coefficients W_k depend on the amplitudes $a_1, \cdots, a_n$ only; they are $\geqq 0$ and their sum equals 1, *so that* (7) *is a certain average of the individual frequencies λ_k. More precisely, W_k is the probability*

$$W_k = W(a_k; a_{k+1} \cdots a_n\, a_1 \cdots a_{k-1}) \tag{8}$$

that an $(n-1)$-linkage with the given sides $a_{k+1} \cdots a_n\, a_1 \cdots a_{k-1}$ spans a distance $< a_k$.

For the proof we first return to our old hypothesis that the λ_k satisfy no linear homogeneous integral relation whatsoever, and then follow a method recently suggested by P. Hartmann, E. R. van Kampen, and A. Wintner [5]. We verify at once that the time derivative

$$\phi'(t) = \Re\left(\frac{1}{2\pi i}\frac{z'}{z}\right) \qquad \{\Re = \text{real part}\}$$

arises by the substitution (2) from the following phase function:

$$\Phi((\vartheta)) = \Re\,\frac{\Sigma\lambda_k a_k e(\vartheta_k)}{\Sigma a_k e(\vartheta_k)} \tag{9}$$

which becomes dangerously infinite on the singular manifold. Were one allowed to apply to this function the principle of equidistribution (6) holding for bounded Riemann integrable functions f, one would at once obtain a mean motion

$$\mu = \int_0^1 \cdots \int_0^1 \Phi((\vartheta))\, d\vartheta_1 \cdots d\vartheta_n. \tag{10}$$

The μ as defined by this formula depends linearly on the λ_k, (7), and its coefficients W_k are integrals of the following type:

$$W_n = \int_0^1 \cdots \int_0^1 \Re \left\{ \frac{a_n e(\vartheta_n)}{\sum_k a_k e(\vartheta_k)} \right\} d\vartheta_1 \cdots d\vartheta_n.$$

So far HKW. I shall now show how to carry out the integration by ϑ_n in W_n.

Indeed, for fixed $\vartheta_1, \cdots, \vartheta_{n-1}$ and variable $\vartheta_n = \vartheta$ we set

$$a_n = a, \quad \sum_{k=1}^{n-1} a_k e(\vartheta_k) = b,$$
$$z = z(\vartheta) = b + a \cdot e(\vartheta).$$

The integrand is then

$$\Re \left(\frac{1}{2\pi i} \frac{z_\vartheta}{z} \right) \tag{11}$$

where the index ϑ designates derivation by ϑ. The integral of (11) by ϑ is $\int d\phi$ where ϕ denotes the direction of the radius vector from the origin O to the variable point $z(\vartheta)$ describing the circle of radius a around the center b, and hence we obtain 1 or 0 according as the origin lies within or without the circle. This distinction amounts to

$$| b | = | a_1 e(\vartheta_1) + \cdots + a_{n-1} e(\vartheta_{n-1}) | \lesseqgtr a_n. \tag{12}$$

We thus find

$$W_n = \int d\vartheta_1 \cdots d\vartheta_{n-1},$$

the integral extending over that part of the $(n-1)$-dimensional phase torus $(\vartheta_1, \cdots, \vartheta_{n-1})$ where

$$| a_1 e(\vartheta_1) + \cdots + a_{n-1} e(\vartheta_{n-1}) | < a_n,$$

or

$$W_n = W(a_n; a_1 \cdots a_{n-1}). \tag{13}$$

The formula (10) for $\Sigma W_k \lambda_k$ yields at once the equation

$$W_1 + \cdots + W_n = 1 \tag{14}$$

by taking $\lambda_1 = \cdots = \lambda_n = 1$. This result has its own interest independent of the problem of mean motion:

Given n positive lengths a_k, let W_k be the probability that an $(n-1)$-linkage with the sides $a_{k+1} \cdots a_n a_1 \cdots a_{k-1}$ spans a distance $< a_k$. Then the n probabilities $W_1, \cdots, W_n$ yield the sum 1.

Certain examples suggest that the equidistribution law (6) might not be applicable to functions with the type of singularity present in $\phi((\vartheta))$. HKW get around this difficulty by resorting to Birkhoff's general ergodic theorem. But then the formula is proved only with the exception of an unknown zero set in the space of the initial phases α_k. Our evaluation of the integral indicates, however, that the evil-spelling singularities of the integrand (9) are more or less bluff. For a complete proof of our theorem I therefore fall back on a much earlier and more elementary procedure inaugurated in 1909 by P. Bohl [6] for the lowest non-trivial case $n = 3$.

We slit the phase torus in the $(n-1)$-dimensional manifold of those points (ϑ) for which $z = z(\vartheta_1 \cdots \vartheta_n)$ is *real and negative*. On the slit torus the azimuth $\phi(\vartheta_1 \cdots \vartheta_n)$ is a continuous (periodic) single-valued function whose values lie between $-\frac{1}{2}$ and $+\frac{1}{2}$ (limits excluded). The boundary of the slit is the singular manifold. On the slit itself ϕ has a discontinuity, taking on the value $+\frac{1}{2}$ on the one, the $+$ side, and $-\frac{1}{2}$ on the $-$ side; it jumps by -1 or $+1$ according as one crosses the slit in the positive or negative sense (i. e. from the $+$ to the $-$ side, or in the opposite direction). Therefore $\phi(t_2) - \phi(t_1)$ differs by a term of absolute value $< \frac{1}{2}$ from the number of times $N(t_1 t_2)$ the straight line $\vartheta(t)$ crosses the slit during the time interval $t_1 \leqq t \leqq t_2$. In computing $N(t_1 t_2)$ the sense of crossing has to be taken into account. Let us erect a cylindrical trunk T over the slit as basis whose generator is the velocity vector $\lambda = (\lambda_1, \cdots, \lambda_n)$. Whenever our moving point $\vartheta(t)$ crosses the slit it will stay in T during the following unit of time, so that $N(t_1 t_2)$ is essentially the duration of its stay in T between t_1 and t_2. Hence $N(t_1 t_2)/(t_2 - t_1)$ will tend to the volume μ of T with $t_2 - t_1 \to \infty$. In computing that volume the multiplicity and orientation in which T covers the various parts of the torus are to be taken into account; the "characteristic function" of the trunk, to which the law of equidistribution (6) is here applied, is the covering index capable of the values $0, \pm 1, \pm 2, \cdots$. μ is perhaps more fittingly described as the flux of the constant velocity field λ through the slit, and is given by the integral over the slit of

$$(15) \qquad \begin{vmatrix} \lambda_1, & \cdots, & \lambda_n \\ d_1\vartheta_1, & \cdots, & d_1\vartheta_n \\ \cdot & \cdot & \cdot \\ d_{n-1}\vartheta_1, & \cdots, & d_{n-1}\vartheta_n \end{vmatrix} .$$

$d_1, \cdots, d_{n-1}$ denote $n-1$ tangential line elements at a point (ϑ) of the slit in such orientation that (15) is positive or negative according as the current λ crosses the slit at (ϑ) in the positive or negative sense. We thus obtain

$$\mu = W_1\lambda_1 + \cdots + W_n\lambda_n \tag{7}$$

where W_n is the area

$$W_n = \int \cdots \int d\vartheta_1 \cdots d\vartheta_{n-1}$$

of the vertical projection of the slit upon the $(n-1)$-dimensional $(\vartheta_1, \cdots, \vartheta_{n-1})$-torus with the orientation of its covering taken into account.

When we wish to ascertain whether (and with what index) a given point $(\vartheta_1 \cdots \vartheta_{n-1})$ is covered by the projection, we draw the $(n-1)$-linkage $A_0A_1 \cdots A_{n-1}$ with the links (3) starting at $A_0 = O$. Denoting A_{n-1} by b we have the figure as described before: the circle of radius $a_n = a$ around b and the origin O. We now add the negative real axis I issuing from O, and we have to watch whether the point $z(\vartheta)$ crosses I as ϑ varies from 0 to 1. $z(\vartheta)$ will cross once and in the positive sense if $|b| < a$; it will not cross at all, or once cross and then recross if $|b| > a$. Thus the covering index $= 1$ or 0 according to the distinction (12), and our previous formula (13) is confirmed.

When looked at in this way the topological situation is as plain as it could be. The flux of a constant velocity field through an $(n-1)$-dimensional surface bounded by an $(n-2)$-dimensional cycle does not depend on what two-sided surface one spans into the cycle; it may be described as the flux encompassed by the cycle and expressed by an integral extending over the cycle. If we had proceeded in this way we would have got entangled in a cobweb of topological difficulties. The singular manifold is of a complex topological structure which, moreover, varies with the values of the parameters a_k; the topologically different cases seem to be separated by those "exceptional" closed n-linkages whose a_k fall apart into two groups of equal sum. The unravelling of this maze is avoided by using the slit rather than the singular manifold, and by defining the slit precisely as we did.

By their very definition as probabilities the W_k are positive. But since we no longer make use of the Wintner integral (10) an independent proof of (14) becomes desirable. From the definition of $z(\vartheta_1 \cdots \vartheta_n)$ it is obvious that

$$\phi(\vartheta_1 + \vartheta, \cdots, \vartheta_n + \vartheta) = \vartheta + \phi(\vartheta_1, \cdots, \vartheta_n) \qquad (\vartheta \text{ arbitrary}),$$

and hence

$$\mu = \Sigma W_k\lambda_k$$

must increase by λ when all frequencies λ_k are raised by the same λ, or $\Sigma W_k = 1$. In arguing thus, one is a little bit encumbered by the assumption that the λ_k are linearly independent. But by choosing $\vartheta = -\vartheta_n$,

$$\phi(\vartheta_1, \cdots, \vartheta_n) = \phi(\vartheta_1 - \vartheta_n, \cdots, \vartheta_{n-1} - \vartheta_n, 0) + \vartheta_n,$$

one carries over the whole discussion from the n-dimensional to an $(n-1)$-dimensional phase space and finds the linear independence of $\lambda_1 - \lambda_n, \cdots, \lambda_{n-1} - \lambda_n$ to be a sufficient condition for our result. But this amounts to the absence of any relation (5) with integral coefficients m_k *of sum* 0, as stated in our theorem, and this hypothesis is not affected by passing from the frequencies λ_k to $\lambda_k + \lambda$.

So far we have assumed that the planet $z(t)$ does not pass through the origin. Should this happen at some moment t_0 the continuation of $\phi(t)$ beyond t_0 may become (and will in general become) ambiguous, with $\phi(t)$ splitting into two equally admissible branches which differ by 1. But our procedure shows that such an occurrence with its attendant accumulating ambiguity is rare enough to be of no influence upon the asymptotic law.

Combining the result (13) with (4) one gets for $n \geqq 3$ an explicit expression in terms of Bessel functions:

$$W_n = a_n \int_0^\infty J_1(a_n\rho) \prod_{k=1}^{n-1} J_0(a_k\rho)\, d\rho. \tag{16}$$

I conclude with a few remarks on the history of our question. When Lagrange hit upon it in his approximate theory of secular perturbations, he at once solved the "trivial" case where one of the sides, a_n say, is larger than the sum of all others,

$$a_n > a_1 + \cdots + a_{n-1}.$$

One then has $\mu = \lambda_n$ with the much sharper estimate

$$\phi(t) = \lambda_n t + O(1).$$

(This estimate was implied in the term "mean motion," as Lagrange used it. Actually all planets are in the Lagrangean case except Venus and Earth.) More than a century elapsed before P. Bohl succeeded in establishing mean motion for $n = 3$ in the non-Lagrangean case where a_1, a_2, a_3 form a triangle Δ. He found

$$\mu = W_1\lambda_1 + W_2\lambda_2 + W_3\lambda_3,$$

πW_k being the angles of Δ. This result is easily obtained if after the aforementioned elementary reduction of the dimensionality of the phase space to $n-1$, one expresses the flux in terms of the singular manifold which now is a 0-cycle. When the problem and Bohl's paper were pointed out to me by Felix Bernstein in 1913, it started me on my investigations on Diophantine approximations (W_1). Although the essential step was accomplished by the equidistribution law, the messy topology of the singular manifold at that time

prevented me from settling the problem by an explicit and universal formula. As an illustration I carried out the case $n = 4$; the topology of closed 4-linkages has been thoroughly investigated in kinematics because such linkages are a frequent element in the construction of machines. Even there I limited myself to one of the topological subcases, see W_1, Satz 7, and blundered in the final formulation (although the proof is correct): since my "integral invariants" of the rickety quadrilateral have the sum 0 instead of 1, the term λ_1 should be added in the final formula. But I will not mar the simplicity of our present argument by these old vagaries. Awareness of those complications will, however, help the reader to a fuller appreciation of the result (8), (16) holding for all cases alike.

This formula, which served me as a lodestar, was guessed for all n and verified for $n = 3, 4$ by Professor Wintner. His report on the paper HKW in my Princeton seminar on current literature stimulated me to resume the ancient problem. The various equations known about Bessel integrals of the type (16) for lower n are most naturally derived from the interpretation put upon (16) by the problem of mean motion.

THE INSTITUTE FOR ADVANCED STUDY,
PRINCETON, N. J.

BIBLIOGRAPHY.

[1] See e. g. Charlier, *Die Mechanik des Himmels* I, Leipzig, 1902, pp. 333-436, and for a brief account the paper W_2 by the author, *Enseignement Mathématique*, vol. 16 (1914), p. 455.
[2] A. Wintner, *American Journal of Mathematics*, vol. 55 (1933), p. 309.
[3] Cf. Watson's *Treatise on Bessel Functions*, pp. 411, 414, and 420.
[4] See the paper W_1 by the author, *Mathematische Annalen*, vol. 77 (1916), p. 313.
[5] *American Journal of Mathematics*, vol. 59 (1937), p. 261, cited as HKW.
[6] P. Bohl, *Journ. reine u. angew. Mathem.*, vol. 135 (1909), p. 189.

THE HOMOGENEOUS CHAOS.

By Norbert Wiener.

---

1. Introduction. Physical need for theory.
2. Definition. Types of chaos.
3. Classical ergodic theorem. Lebesgue form.
4. Dominated ergodic theorem. Multidimensional ergodic theorem.
5. Metric transitivity. Space and phase averages in a chaos.
6. Pure one-dimensional chaos.
7. Pure multidimensional chaos.
8. Phase averages in a pure chaos.
9. Forms of chaos derivable from a pure chaos.
10. Chaos theory and spectra.
11. The discrete chaos.
12. The weak approximation theorem for the polynomial chaos.
13. The physical problem. The transformations of a chaos.

1. Introduction. Physical need for theory. Statistical mechanics may be defined as the application of the concepts of Lebesgue integration to mechanics. Historically, this is perhaps putting the cart before the horse. Statistical mechanics developed through the entire latter half of the nineteenth century before the Lebesgue integral was discovered. Nevertheless, it developed without an adequate armory of concepts and mathematical technique, which is only now in the process of development at the hands of the modern school of students of integral theory.

In the more primitive forms of statistical mechanics, the integration or summation was taken over the manifold particles of a single homogeneous dynamical system, as in the case of the perfect gas. In its more mature form, due to Gibbs, the integration is performed over a parameter of distribution, numerical or not, serving to label the constituent systems of a dynamical ensemble, evolving under identical laws of force, but differing in their initial conditions. Nevertheless, the study of the mode in which this parameter of distribution enters into the individual systems of the ensemble does not seem to have received much explicit study. The parameter of distribution is essentially a parameter of integration only. As such, questions of dimensionality are indifferent to it, and it may be replaced by a numerical variable α with the

range $(0, 1)$. Any transformation leaving invariant the probability properties of the ensemble as a whole is then represented by a measure-preserving transformation of the interval $(0, 1)$ into itself.

Among the simplest and most important ensembles of physics are those which have a spatially homogeneous character. Among these are the homogeneous gas, the homogeneous liquid, the homogeneous state of turbulence. In these, while the individual systems may not be invariant under a change of origin, or, in other words, under the translation of space by a vector, the ensemble as a whole is invariant, and the individual systems are merely permuted without change of probability. From what we have said, the translations of space thus generate an Abelian group of equi-measure transformations of the parameter of distribution.

One-dimensional groups of equi-measure transformations have become well known to the mathematicians during the past decade, as they lie at the root of Birkhoff's famous ergodic theorem.[1] This theorem asserts that if we have given a set S of finite measure, an integrable function $f(P)$ on S, and a one-parameter Abelian group T^λ of equi-measure transformations of S into itself, such that

$$T^\lambda \cdot T^\mu = T^{\lambda+\mu} \qquad (-\infty < \lambda < \infty,\ -\infty < \mu < \infty), \tag{1}$$

then for all points P on S except those of a set of zero measure, and provided certain conditions of measurability are satisfied,

$$\lim_{A\to\infty} \frac{1}{A} \int_0^A f(T^\lambda P)\, d\lambda \tag{2}$$

will exist. Under certain more stringent conditions, known as metric transitivity, we shall have

$$\lim_{A\to\infty} \frac{1}{A} \int_0^A f(T^\lambda P)\, d\lambda = \int_S f(P)\, dV_P \tag{3}$$

almost everywhere. The ergodic theorem thus translates averages over an infinite range, taken with respect to λ, into averages over the set S of finite measure. Even without metric transitivity, the ergodic theorem translates the distribution theory of λ averages into the theory of S averages.

In the most familiar applications of the ergodic theorem, S is taken to be a spatial set, and the parameter λ is identified with the time. The theorem thus becomes a way of translating time averages into space averages, in a

[1] Cf. Eberhard Hopf, "Ergodentheorie," *Ergebnisse der Mathematik und ihrer Grenzgebiete*, vol. 5. See particularly § 14, where further references to the literature are given.

manner which was postulated by Gibbs without rigorous justification, and which forms the entire basis of his methods. Strictly speaking, the space averages are generally in phase-space rather than in the ordinary geometrical space of three dimensions. There is no reason, however, why the parameter λ should be confined to one taking on values of the time for its arguments, nor even why it should be a one-dimensional variable. We are thus driven to formulate and prove a multidimensional analogue of the classical Birkhoff theorem.

In the ordinary Birkhoff theorem, the transformations T^λ are taken to be one-one point transformations. Now, the ergodic theorem belongs fundamentally to the abstract theory of the Lebesgue integral, and in this theory, individual points play no rôle. In the study of chaos, individual values of the parameter of integration are equally unnatural as an object of study, and it becomes desirable to recast the ergodic theorem into a true Lebesgue form. This we do in paragraph 2.

Of all the forms of chaos occurring in physics, there is only one class which has been studied with anything approaching completeness. This is the class of types of chaos connected with the theory of the Brownian motion. In this one-dimensional theory, there is a simple and powerful algorithm of phase averages, which the ergodic theorem readily converts into a theory of averages over the transformation group. This theory is easily generalized to spaces of a higher dimensionality, without any very fundamental alterations. We shall show that there is a certain sense in which these types of chaos are central in the theory, and allow us to approximate to all types.

Physical theories of chaos, such as that of turbulence, or of the statistical theory of a gas or a liquid, may or may not be theories of equilibrium. In the general case, the statistical state of a chaotic system, subject to the laws of dynamics, will be a function of the time. The laws of dynamics produce a continuous transformation group, in which the chaos remains a chaos, but changes its character. This is at least the case in those systems which can continue to exist indefinitely in time without some catastrophe which essentially changes their dynamic character. The study, for example, of the development of a state of turbulence, depends on an existence theory which avoids the possibility of such a catastrophe. We shall close this paper by certain very general considerations concerning the demands of an existence theory of this sort.

2. Definition. Types of Chaos. A *continuous homogeneous chaos* in n dimensions is a scalar or vector valued measurable function $\rho(x_1, \cdots, x_n; \alpha)$ of $x_1, \cdots, x_n; \alpha$, in which $x_1, \cdots, x_n$ assume all real values, while α ranges over $(0, 1)$; and in which the set of values of α for which

(4) $\rho(x_1 + y_1, \cdots, x_n + y_n; \alpha)$ belongs to S,

if it has a measure for any set of values of $y_1, \cdots, y_n$, has the same measure for any other set of values. In this paper, we shall confine our attention to scalar chaoses. A continuous homogeneous chaos is said to be *metrically transitive*, if whenever the sets of values of α for which $\rho(x_1, \cdots, x_n; \alpha)$ belongs to S and to S_1, respectively, have measures M and M_1, the set of values of α for which simultaneously

$$\rho(x_1, \cdots, x_n; \alpha) \text{ belongs to } S$$

and

$$\rho(x_1 + y_1, \cdots, x_n + y_n; \alpha) \text{ belongs to } S_1$$

has a measure which tends to M_1 as $y_1^2 + \cdots + y_n^2 \to \infty$.

If ρ is integrable, it determines the additive set-function

$$(5) \qquad \mathfrak{F}(\Sigma; \alpha) = \int \cdots \int_{\Sigma} \rho(x_1, \cdots, x_n; \alpha)\, dx_1 \cdots dx_n.$$

On the other hand, not every additive set-function may be so defined. This suggests a more general definition of a *homogeneous chaos*, in which the chaos is defined to be a function $\mathfrak{F}(\Sigma; \alpha)$, where α ranges over $(0, 1)$ and Σ belongs to some additively closed set Ξ of measurable sets of points in n-space. We suppose that if Σ and Σ_1 do not overlap,

$$(6) \qquad \mathfrak{F}(\Sigma + \Sigma_1; \alpha) = \mathfrak{F}(\Sigma; \alpha) + \mathfrak{F}(\Sigma_1; \alpha).$$

We now define the new point-set $\Sigma(y_1, \cdots, y_n)$ by the assertion

(7) $\Sigma(y_1, \cdots, y_n)$ contains $x_1 + y_1, \cdots, x_n + y_n$ when and only when Σ contains $x_1, \cdots, x_n$.

This leads to the definition of the additive set-function $\mathfrak{F}_{y_1, \ldots, y_n}(\Sigma; \alpha)$ by

$$(8) \qquad \mathfrak{F}_{y_1, \ldots, y_n}(\Sigma; \alpha) = \mathfrak{F}(\Sigma(y_1, \cdots, y_n); \alpha).$$

If, then, for all classes S of real numbers,

(9) Measure of set of α's for which $\mathfrak{F}_{y_1, \ldots, y_n}(\Sigma; \alpha)$ belongs to class S

is independent of $y_1, \cdots, y_n$, in the sense that if it exists for one set of these numbers, it exists for all sets, and has the same value, and if it is measurable in $y_1, \cdots, y_n$, we shall call $\mathfrak{F}$ a homogeneous chaos. The notion of metrical transitivity is generalized in the obvious way, replacing $\rho(x_1, \cdots, x_n; \alpha)$ by $\mathfrak{F}(\Sigma; \alpha)$, and $\rho(x_1 + y_1, \cdots, x_n + y_n; \alpha)$ by $\mathfrak{F}_{y_1, \ldots, y_n}(\Sigma; \alpha)$.

The theorem which we wish to prove is the following:

THEOREM I. *Let* $\mathfrak{F}(\Sigma;\alpha)$ *be a homogeneous chaos. Let the functional*

$$\Phi\{\mathfrak{F}(\Sigma;\alpha)\} = g(\alpha) \tag{10}$$

be a measurable function of α, *such that* $\int_0^1 |\, g(\alpha) \log^+ |\, g(\alpha)|\,|\, d\alpha$ *is finite. Then for almost all values of* α,

$$\lim_{r\to\infty} \frac{1}{V(r)} \int \cdots \int_R \Phi\{\mathfrak{F}_{y_1,\ldots,y_n}(\Sigma;\alpha)\}dy_1 \cdots dy_n \tag{11}$$

exists, where R *is the interior of the sphere*

$$y_1^2 + y_2^2 + \cdots + y_n^2 = r^2 \tag{12}$$

and $V(r)$ *is its volume. If in addition,* $\mathfrak{F}(\Sigma;\alpha)$ *is metrically transitive,*

$$\lim_{r\to\infty} \frac{1}{V(r)} \int \cdots \int_R \Phi\{\mathfrak{F}_{y_1,\ldots,y_n}(\Sigma;\alpha)\}dy_1 \cdots dy_n = \int_0^1 \Phi\{\mathfrak{F}(\Sigma;\beta)\}d\beta \tag{13}$$

for almost all values of α.

3. Classical ergodic theorem. Lebesgue form. Theorem I is manifestly a theorem of the ergodic type. Let it be noticed, however, that we nowhere assume that the transformation of α given by $\beta = T\alpha$ when

$$\mathfrak{F}_{y_1,\ldots,y_n}(\Sigma;\beta) = \mathfrak{F}(\Sigma;\alpha) \tag{14}$$

is one-one. This should not be surprising, as the ergodic theorem is fundamentally one concerning the Lebesgue integral, and in the theory of the Lebesgue integral, individual points play no rôle.

Nevertheless, in the usual formulation of the ergodic theorem, the expression $f(T^\lambda P)$ enters in an essential way. Can we give this a meaning without introducing the individual transform of an individual point?

The clue to this lies in the definition of the Lebesgue integral itself. If $f(P)$ is to be integrated over a region S, we divide S into the regions $S_{a,b}(f)$, defined by the condition that over such a region,

$$a < f(P) \leq b. \tag{15}$$

We now write

$$\int_S f(P)\,dV_P = \lim_{\epsilon\to 0} \sum_{-\infty}^{\infty} n\,\epsilon\, m(S_{(n-1)\epsilon,\, n\epsilon}(f)). \tag{16}$$

The condition that $f(P)$ be integrable thus implies the condition that it be measurable, or that all the sets $S_{a,b}(f)$ be measurable.

Now, if T is a measure-preserving transformation on S, the sets $TS_{a,b}(f)$

will all be measurable, and will have, respectively, the same measures as the sets $S_{a,b}(f)$. We shall define the function $f(TP) = g(P)$ by the conditions

$$TS_{a,b}(f) = S_{a,b}(g). \tag{17}$$

If T conserves relations of inclusion of sets, up to sets of zero measure, this function will clearly be defined up to a set of values of P of zero measure, and we shall have

$$\int_S f(TP)\,dV_P = \int_S f(P)\,dV_P. \tag{18}$$

We may thus formulate the original or discrete case of the Birkhoff ergodic theorem, as follows: *Let S be a set of points of finite measure. Let T be a transformation of all measurable sub-sets of S into measurable sub-sets of S, which conserves measure, and the relation between two sets, that one contains the other except at most for a set of zero measure. Then except for a set of points P of zero measure,*

$$\lim_{N\to\infty} \frac{1}{N+1} \sum_0^N f(T^n P) \tag{19}$$

will exist.

The continuous analogue of this theorem needs to be formulated in a somewhat more restricted form, owing to the need of providing for the integrability of the functions concerned. It reads: *Let S be a set of points of finite measure. Let T^λ be a group of transformations fulfilling the conditions we have laid down for T in the discrete case just mentioned. Let $T^\lambda P$ be measurable in the product space of λ and of P. Then, except for a set of points P of zero measure,*

$$\lim_{A\to\infty} \frac{1}{A} \int_0^A f(T^\lambda P)\,d\lambda \tag{20}$$

will exist.

In the proofs of Birkhoff's ergodic theorem, as given by Khintchine and Hopf, no actual use is made of the fact that the transformation T is one-one, and the proofs extend to our theorem as stated here, without any change. The restriction of measurability, or something to take its place, is really necessary for the correct formulation of Khintchine's statement of the ergodic theorem, as Rademacher, von Neumann, and others have already pointed out.[2]

With the aid of the proper Lebesgue formulation of the ergodic theorem, the one-dimensional case of Theorem I follows at once. Actually more follows, as it is only necessary that g belong to L, instead of to the logarithmic class

[2] Cf. J. v. Neumann, *Annals of Mathematics*, 2, vol. 33, p. 589, note 11.

with which we replace it. To prove Theorem I in its full generality, we must establish a multidimensional ergodic theorem.

4. Dominated ergodic theorem. Multidimensional ergodic theorem. As a lemma to the multidimensional ergodic theorem, we first wish to establish the fact that if the function $f(P)$ in the ergodic theorem satisfies the condition

$$\int_S |f(P)| \log^+ |f(P)| \, dV_P < \infty, \tag{21}$$

then the expressions (19) and (20) not only exist, but the limits in question will be approached under the domination of a summable function of P. We shall prove this in the discrete case, for the sake of simplicity, but the result goes over without difficulty to the continuous case.

Let T be an equimeasure transformation of the set S of finite measure into itself, in the generalized sense of the last paragraph, and let W be a measurable sub-set of S, with the characteristic function $W(P)$. Let U be the set of all points P for which some $T^{-j}P$ belongs to W. Let $i(P)$, when P belongs to W, be defined as the smallest positive number n such that $T^{-n}P$ belongs to W, and let us call it the *index* of P. Every point of W, except for a set of measure 0, will have a finite index, since if we write W_∞ for the set of points without a finite index, no two sets $T^m W_\infty$ and $T^n W_\infty$ can overlap, while they all have the same measure and their sum has a finite measure. Thus except for a set of zero measure, we may divide W into the sets W_p, each consisting of the points of W of index p. It is easy to show that if $1 \leq p < \infty$, $1 \leq p' < \infty$, $0 \leq j < p$, $0 \leq j' < p'$, the sets $T^{-j}W_p$ and $T^{-j'}W_{p'}$ can not overlap over a set of positive measure unless $j = j'$, $p = p'$. Similarly, the sets $T^{-p}W_p$ and $T^{-p'}W_{p'}$ can not overlap over sets of positive measure, unless $p = p'$, and represent a dissection of W, except for a set of zero measure. Let us put W_{pq} for the logical product of W_p and $T^p W_q$; W_{pqr} for the logical product of W_{pq} and $T^{p+q}W_r$; and so on. Then if k is fixed, the sets $T^{-j}W_{p_1 p_2 \dots p_k}$ cover U (except for sets of zero measure) once as j goes from 0 to p_1, once as it goes from p_1 to $p_1 + p_2$, and so on; making just k times between 0, inclusive, and $p_1 + p_2 + \cdots + p_k$, exclusive. Thus if S_K is the set of all the points in all the $W_{p_1 p_2 \dots p_k}$ for which $p_1 + p_2 + \cdots + p_k = K$, the total measure of all the S_K's for $2^{N+1} \geq K \geq 2^N$ can not exceed $km(S)/2^N$.

Now let P lie in $T^{-j}W_{p_1 p_2 \dots p_k}$, where $0 \leq j < p_1 + p_2 + \cdots + p_k$. Let us consider the sequence of numbers a_j, where if $p_1 + \cdots + p_l \leq j < p_1 + \cdots + p_{l+1}$, a_j is the greatest of the numbers

$$\frac{1}{j - p_1 - \cdots - p_l + 1}, \quad \frac{2}{j - p_1 - \cdots - p_{l-1} + 1}, \quad \frac{3}{j - p_1 - \cdots - p_{l-2} + 1}, \cdots, \frac{l+1}{j+1}.$$

Then a_j will be the largest of the numbers

$$W(P), \frac{W(P)+W(TP)}{2}, \cdots, \frac{W(P)+W(TP)+\cdots+W(T^jP)}{j+1}.$$

The sum $\sum_0^k a_j$, for a fixed K, will have its maximum value when $p_1 = p_2 = \cdots = p_{k-1} = 1$; $p_k = K+1-k$, when it will be $k + \sum_{k+1}^{k} k/j \leq k(1+\log K/k)$. In this case the sequence of the a_j's will be

$$\underbrace{1, 1, \cdots, 1}_{k \text{ times}}, \frac{k}{k+1}, \frac{k}{k+2}, \cdots, \frac{k}{K}.$$

This remark will be an easy consequence of the following fact: let us consider the sequence

$$(22) \qquad \cdots, \frac{\lambda}{n-1}, \frac{\lambda}{n}, 1, \frac{1}{2}, \cdots, \frac{1}{\left[\frac{n}{\lambda}\right]}, \frac{\lambda+1}{\left[\frac{n}{\lambda}\right]+n+1}, \cdots$$

and the modified sequence

$$(23) \qquad \cdots, \frac{\lambda}{n-1}, 1, \frac{1}{2}, \cdots, \frac{1}{\left[\frac{n-1}{\lambda}\right]}, \frac{\lambda+1}{\left[\frac{n-1}{\lambda}\right]+n}, \frac{\lambda+1}{\left[\frac{n-1}{\lambda}\right]+n+1}, \cdots$$

where of course

$$\frac{\lambda+1}{n+\left[\frac{n}{\lambda}\right]} \leq \frac{1}{\left[\frac{n}{\lambda}\right]}; \quad \frac{\lambda+1}{n+\left[\frac{n}{\lambda}\right]+1} > \frac{1}{\left[\frac{n}{\lambda}\right]+1};$$

$$\frac{\lambda+1}{n-1+\left[\frac{n-1}{\lambda}\right]} \leq \frac{1}{\left[\frac{n-1}{\lambda}\right]}; \quad \frac{\lambda+1}{n+\left[\frac{n-1}{\lambda}\right]} > \frac{1}{\left[\frac{n-1}{\lambda}\right]+1}.$$

Apart from the arrangement, the terms of (22) will be the same as the terms of (23), except that in (23), $\dfrac{\lambda+1}{\left[\frac{n-1}{\lambda}\right]+n}$ replaces λ/n. Now,

$$\frac{\lambda+1}{\left[\frac{n-1}{\lambda}\right]+n} > \frac{\lambda}{n},$$

so that the sum of the terms in (23) is greater than that in (22).

Since the transforms of a given set have the same measure as the set, and the sets $T^{-j}S_K$ cover U exactly k times, we have

$$\begin{aligned}
&\frac{1}{k}\int_S \Big\{ W(P) + \max\left(W(P), \frac{W(P)+W(TP)}{2}\right) \\
&\quad + \max\left(W(P), \frac{W(P)+W(TP)}{2}, \frac{W(P)+W(TP)+W(T^2P)}{3}\right) \\
&\quad + \cdots + \max\left(W(P), \frac{W(P)+W(TP)}{2}, \cdots, \right. \\
&\qquad\qquad \left. \frac{W(P)+\cdots+W(T^{k-1}P)}{k}\right)\Big\}\, dV_P \\
&\leq \sum_{K=0}^{\infty} m(S_K)\left(1+\log\frac{K}{k}\right) \\
&\leq m(W)\left(1+\log\frac{2^N}{k}\right) + \sum_{j=1}^{\infty}\frac{km(S)}{2^{N+j}}\left(1+\log\frac{2^{N+j}}{k}\right) \\
&\leq \left(1+\log\frac{2^N}{k}\right)\left\{ m(W) + \text{const.}\,\frac{km(S)}{2^N}\right\},
\end{aligned}$$

the constant being absolute. If we now put

$$N = \left[\log\frac{km(S)}{m(W)} \Big/ \log 2\right],$$

this last expression is dominated by

$$\text{const.}\, m(W)\left(1+\log\frac{m(S)}{m(W)}\right).$$

Since the constant is independent of k, we see that the Cesàro average of

$$\cdot \max\left(W(P), \frac{W(P)+W(TP)}{2}, \cdots, \frac{W(P)+\cdots+W(T^{k-1}P)}{k}\right)$$

is dominated by the same term. Thus

$$\begin{aligned}
&\int_S \max\left(W(P), \cdots, \frac{W(P)+\cdots+W(T^{k-1}P)}{k}\right) dV_P \\
&\qquad \leq \text{const.}\, m(W)\left(1+\log\frac{m(S)}{m(W)}\right),
\end{aligned}$$

and it follows by monotone convergence that there exists a function $W^*(P)$ such that

$$\int_S W^*(P)\, dV_P \leq \text{const.}\, m(W)\left(1+\log\frac{m(S)}{m(W)}\right), \tag{24}$$

and for all positive m,

$$\frac{W(P)+\cdots+W(T^mP)}{m+1} \leq W^*(P). \tag{25}$$

Now let $f(P)$ be a function such that

$$\int_S |f(P)| \log^+ |f(P)| \, dV_P < \infty. \tag{21}$$

Let $W^{(N)}$ be the set of points such that

$$2^N \leq |f(P)| \leq 2^{N+1}, \tag{26}$$

and let $f^{(N)}(P)$ be the function equal to $f(P)$ over this set of points, and 0 elsewhere. Then there exists a function $f^{(N)*}(P)$, such that

$$\frac{f^{(N)}(P) + \cdots + f^{(N)}(T^m P)}{m+1} \leq f^{(N)*}(P) \quad (m = 0, 1, 2, \cdots), \tag{27}$$

and

$$\int_S f^{(N)*}(P) dV_P \leq \text{const. } m(W^{(N)}) \left(1 + \log^+ \frac{m(S)}{m(W^{(N)})}\right) 2^N. \tag{28}$$

Hence if

$$\sum_{-\infty}^{\infty} m(W^{(N)}) \left(1 + \log^+ \frac{m(S)}{m(W^{(N)})}\right) 2^N < \infty \tag{29}$$

and

$$f^*(P) = \sum_{-\infty}^{\infty} f^{(N)*}(P), \tag{30}$$

then

$$\frac{f(P) + \cdots + f(T^m P)}{m+1} \leq f^*(P) \qquad (m = 0, 1, 2, \cdots), \tag{31}$$

and

$$\int_S f^*(P) dV_P \leq \text{const. } \Sigma \, m(W^{(N)}) \left(1 + \log^+ \frac{m(S)}{m(W^{(N)})}\right) 2^N. \tag{32}$$

However,

$$\begin{aligned} &\sum_{-\infty}^{\infty} m(W^{(N)}) \left(1 + \log^+ \frac{m(W^{(N)})}{m(S)}\right) 2^N \\ &\leq \sum_{-\infty}^{\infty} \int_{W^{(N)}} |f(P)| \, dV_P \left(1 + \log^+ \frac{2^{N+1} m(S)}{\int_S |f(P)| \, dV_P}\right) \\ &\leq \int_S |f(P) \left(1 + \log^+ \frac{2 |f(P)| \, m(S)}{\int_S |f(Q) dV_Q}\right) dV_P \\ &\leq \int_S |f(P)| \, dV_P \left(1 + \log^+ \frac{2m(S)}{\int_S |f(P)| \, dV_P}\right) \\ &\quad + \int_S |f(P)| \log^+ |f(P)| \, dV_P. \end{aligned} \tag{33}$$

Thus $\int_S f^*(P)\,dV_P$ has an upper bound which is less than a function of $\int_S |f(P)|\,dV_P$ and $\int_S |f(P)| \log^+ |f(P)|\,dV_P$, tending to 0 as they both tend to 0. This establishes our theorem of the existence of a uniform dominant.

There is a sense in which (21) *is a best possible condition.* That is, if

$$\psi(x) = o(\log^+ x), \tag{34}$$

the condition

$$\int_S (\psi(|f(P)|)\,|f(P)|\,dV_P < \infty \tag{35}$$

is not sufficient for the existence of a uniform dominant. For let S be a set of measure 1, subdivided into mutually exclusive sets S_n, of measures respectively 2^{-n}. Let S_n be divided into mutually exclusive sets $S_{n,1}, \cdots, S_{n,\nu_n}$, all of equal measures. Let T transform $S_{n,k}$ into $S_{n,(k+1)}$ $(k < \nu^n)$, and S_{n,ν_n} into $S_{n,1}$. Let $f(P)$ be defined by

$$f(P) = a_n > 0 \text{ on } S_{n_1}, \quad (n = 1, 2, \cdots); \quad f(P) = 0 \text{ elsewhere.} \tag{36}$$

Then the smallest possible uniform dominant of

$$\frac{1}{N+1} \sum_0^N f(T^n P) \tag{37}$$

is

$$f^*(P) = \frac{a_n}{k} \text{ on } S_{n,k}, \tag{38}$$

and we have

$$\int_S f^*(P)\,dV_P = \sum \frac{a_n \Omega(\log \nu_n) 2^{-n}}{\nu_n}. \tag{39}$$

Thus if

$$\nu_n = 2^{2n} a_n, \qquad a_n = \Omega(n!), \tag{40}$$

the function $f^*(P)$ will belong to L if and only if

$$\infty > \sum \frac{a_n \Omega(\log a_n) 2^{-n}}{\nu_n} > \text{const.} \int_S f(P) \log^+ f(P)\,dV_P. \tag{41}$$

While we have proved the dominated ergodic theorem merely as a lemma for the multidimensional ergodic theorem, the theorem, and more particularly, the method by which we have proved it, have very considerable independent interest. We may use these methods to deduce von Neumann's mean ergodic theorem from the Birkhoff theorem; or vice versa, we may deduce the Birkhoff theorem, at least in the case of a function satisfying (21), from the von

Neumann theorem. These facts however are not relevant to the frame of the present paper, and will be published elsewhere.

We shall now proceed to the proof of the multidimensional ergodic theorem, which we shall establish in the two-dimensional case, although the method is independent of the number of dimensions. Let $T_1^{\lambda}T_2^{\mu}$ be a two-dimensional Abelian group of transformations of the set S (of measure 1) into itself, in the sense in which we have used this term in paragraph 3. Let $T_1^{\lambda}T_2^{\mu}P$ be measurable in λ, μ, and P. We now introduce a new variable x, ranging over $(0, 1)$, and form the product space $\mathfrak{S}$ of P and x. We introduce the one-parameter group of transformations of this space, T^{ρ}, by putting

$$T^{\rho}(P, x) = (T_1^{\rho \cos 2\pi x} T_2^{\rho \sin 2\pi x} P, x) \tag{42}$$

The expressions $T^{\rho}(P, x)$ will be measurable in ρ and (P, x), and the transformations T^{ρ} will all preserve measure on $\mathfrak{S}$. Thus by the ergodic theorem, for almost all points (P, x) of $\mathfrak{S}$, if $f(P) = f(P, x)$ belongs to L, the limit

$$\lim_{A\to\infty} \frac{1}{A} \int_0^A f(T^{\rho}(P, x))\, d\rho = \lim_{A\to\infty} \frac{1}{A} \int_0^A f(T_1^{\rho \cos 2\pi x} T_2^{\rho \sin 2\pi x} P)\, d\rho \tag{43}$$

will exist. If condition (26) is satisfied, it will follow by dominated convergence that the limit

$$\lim_{A\to\infty} \frac{1}{A} \int_0^A d\rho \int_0^1 f(T_1^{\rho \cos 2\pi x} T_2^{\rho \sin 2\pi x} P)\, dx \tag{44}$$

will exist for almost all points (P, x), and hence for almost all points P.

For the moment, let us assume that $f(P)$ is non-negative. Then there is a Tauberian theorem, due to the author,[3] which establishes that the expression (44) is equivalent to

$$\lim_{A\to\infty} \frac{1}{\pi A^2} \int_0^A \rho d\rho \int_0^{2\pi} f(T_1^{\rho \cos\theta} T_2^{\rho \sin\theta} P)\, d\theta. \tag{45}$$

The only point of importance which we must establish in order to justify this Tauberian theorem is that

$$\int_0^1 \rho^{1+iu} d\rho = \frac{1}{2 + iu} \neq 0 \tag{46}$$

for real values of u. Since every function $f(P)$ satisfying (21) is the difference of two non-negative functions satisfying this condition, (45) is established in the general case.

It will be observed that we have established our multidimensional ergodic

[3] N. Wiener, "Tauberian theorems," *Annals of Mathematics*, 2, vol. 33, p. 28.

theorem on the basis of assumption (21), and not on that of the weaker assumption that $f(P)$ belongs to L. What the actual state of affairs may be, we do not know. At any rate, all attempts to arrive at a direct analogue of the Khintchine proof for one dimension have broken down. The one-dimensional proof makes essential use of the fact that the difference of two intervals is always an interval, while the difference between two spheres is not always a sphere.

The precise statement of the multidimensional ergodic theorem is the following: *Let S be a set of points of measure 1, and let $T_1^{\lambda_1}T_2^{\lambda_2}\cdots T_n^{\lambda_n}$ be an Abelian group of equimeasure transformations of S into itself, in the sense of paragraph 3. Let $T_1^{\lambda_1}\cdots T_n^{\lambda_n}P$ be measurable in $\lambda_1,\cdots,\lambda_n$; P. Let R be the set of values of $\lambda_1,\cdots,\lambda_n$ for which*

$$\lambda_1^2+\lambda_2^2+\cdots+\lambda_n^2\leq r^2 \tag{47}$$

and let $V(r)$ be its volume. Let $f(P)$ satisfy the condition (21). Then for almost all values of P,

$$\lim_{r\to\infty}\frac{1}{V(r)}\int\cdots\int_R f(T_1^{\lambda_1}\cdots T_n^{\lambda_n}P)\,d\lambda_1\cdots d\lambda_n \tag{48}$$

exists.

That part of Theorem I which does not concern metric transitivity is an immediate corollary.

5. Metric transitivity. Space and Phase Averages in a Chaos. If the function $f(P)$ is positive, clearly

$$\begin{aligned}\int\cdots\int_{\lambda_1^2+\ldots+\lambda_n^2\leq r^2-\mu_1^2-\ldots-\mu_n^2} & f(T_1^{\lambda}\cdots T_n^{\lambda}P)\,d\lambda_1\cdots d\lambda_n \\ \leq\int\cdots\int_{\lambda_1^2+\ldots+\lambda_n^2\leq r^2} & f(T_1^{\lambda_1}\cdots T_n^{\lambda_n}T_1^{\mu_1}\cdots T_n^{\mu_n}P)\,d\lambda_1\cdots d\lambda_n \\ \leq\int\cdots\int_{\lambda_1^2+\ldots+\lambda_n^2\leq r^2+\mu_1^2+\ldots+\mu_n^2} & f(T_1^{\lambda_1}\cdots T_n^{\lambda_n}P)\,d\lambda_1\cdots d\lambda_n.\end{aligned} \tag{49}$$

Hence

$$\begin{aligned}\lim_{r\to\infty}\frac{1}{V(r)}\int\cdots\int_R & f(T_1^{\lambda_1}\cdots T_n^{\lambda_n}T_1^{\mu_1}\cdots T_n^{\mu_n}P)\,d\lambda_1\cdots d\lambda_n \\ =\lim_{r\to\infty}\frac{1}{V(r)}\int\cdots\int_R & f(T_1^{\lambda_1}\cdots T_n^{\lambda_n}P)\,d\lambda_1\cdots d\lambda_n,\end{aligned} \tag{50}$$

and expression (48) has the same value for P and all its transforms under the group $T_1^{\lambda_1}\cdots T_n^{\lambda_n}$. The condition of positivity is clearly superfluous. Thus in case expression (48) does not almost everywhere assume a single

value, there will be two classes S_1 and S_2 of elements of S, each of positive measure, and each invariant under all the transformations $T_1^{\lambda_1} \cdots T_n^{\lambda_n}$.

A condition which will manifestly exclude such a contingency is that if S_1 and S_2 are two sub-sets of S of positive measure, and ϵ is a positive quantity, there always exists a transformation $T = T_1^{\lambda_1} \cdots T_n^{\lambda_n}$, such that

$$\left| \frac{mS_1(TS_2)}{mS_2} - mS_1 \right| < \epsilon. \tag{51}$$

From this it will immediately follow that if a chaos is metrically transitive in the sense of paragraph 3, the group of transformations of the α space generated by translations of the chaos will have the property we have just stated, and under the assumptions of Theorem I,

$$\lim_{r\to\infty} \frac{1}{V(r)} \int_R \cdots \int \Phi\{\mathfrak{F}_{y_1, \ldots, y_n}(\Sigma; \alpha)\} dy_1 \cdots dy_n \tag{52}$$

will exist and have the same value for almost all values of α.

If almost everywhere

$$\lim_{r\to\infty} \frac{1}{V(r)} \int_R \cdots \int \Phi\{\mathfrak{F}_{y_1, \ldots, y_n}(\Sigma; \alpha)\} dy_1 \cdots dy_n = A, \tag{53}$$

and

$$\frac{1}{V(r)} \int_R \cdots \int \Phi\{\mathfrak{F}_{y_1, \ldots, y_n}(\Sigma; \alpha)\} dy_1 \cdots dy_n < g(\alpha) \tag{54}$$

where $g(\alpha)$ belongs to L, then by dominated convergence,

$$\begin{aligned} A &= \lim_{r\to\infty} \frac{1}{V(r)} \int_R \cdots \int dy_1 \cdots dy_n \int_0^1 d\alpha \Phi\{\mathfrak{F}_{y_1, \ldots, y_n}(\Sigma; \alpha)\} \\ &= \int_0^1 \Phi\{\mathfrak{F}(\Sigma; \alpha)\} d\alpha. \end{aligned} \tag{55}$$

That is, the average of $\Phi\{\mathfrak{F}(\Sigma; \alpha)\}$, taken over the finite phase space of α, is almost everywhere the same as the average of $\Phi\{\mathfrak{F}_{y_1, \ldots, y_n}(\Sigma; \alpha)\}$ taken over the infinite group space of points $y_1, \cdots, y_n$. This completes the establishment of Theorem I, and gives us a real basis for the study of the homogeneous chaos.[4]

6. Pure one-dimensional chaos. The simplest type of pure chaos is that which has already been treated by the author in connection with the Brownian motion. However, as we wish to generalize this theory to a multi-

[4] The material of this chapter, in the one-dimensional case, has been discussed by the author with Professor Eberhard Hopf several years ago, and he wishes to thank Professor Hopf for suggestions which have contributed to his present point of view.

plicity of dimensions, instead of referring to existing articles on the subject, we shall present it in a form which emphasizes its essential independence of dimensionality.

The type of chaos which we shall consider is that in which the expression $\mathfrak{F}(\Sigma;\alpha)$ has a distribution in α dependent only on the measure of the set Σ; and in which, if Σ_1 and Σ_2 do not overlap, the distributions of $\mathfrak{F}(\Sigma_1,\alpha)$ and $\mathfrak{F}(\Sigma_2,\alpha)$ are independent, in the sense that if $\phi(x,y)$ is a measurable function, and either side of the equation has a sense,

$$(56)\quad \int_0^1\int_0^1 \phi(\mathfrak{F}(\Sigma_1;\alpha),\mathfrak{F}(\Sigma_2,\beta))\,d\alpha d\beta = \int_0^1 \phi(\mathfrak{F}(\Sigma_1,\alpha),\mathfrak{F}(\Sigma_2,\alpha))\,d\alpha.$$

We assume a similar independence when n non-overlapping sets $\Sigma_1,\cdots,\Sigma_n$ are concerned. It is by no means intuitively certain that such a type of chaos exists. In establishing its existence, we encounter a difficulty belonging to many branches of the theory of the Lebesgue integral. The fundamental theorem of Lebesgue assures us of the possibility of adding the measures of a denumerable assemblage of measurable sets, to get the measure of their sum, if they do not overlap. Accordingly, behind any effective realization of the theory of Lebesgue integration, there is always a certain denumerable family of sets in the background, such that all measurable sets may be approximated by denumerable combinations of these. This family is not unique, but without the possibility of finding it, there is no Lebesgue theory.

On the other hand, a theory of measure suitable for the description of a chaos must yield the measure of any assemblage of functions arising from a given measurable assemblage by a translational change of origin. This set of assemblages is essentially non-denumerable. Any attempt to introduce the notion of measure in a way which is invariant under translational changes of origin, without the introduction of some more restricted set of measurable sets, which does not possess this invariance, will fail to establish those essential postulates of the Lebesgue integral which deal with denumerable sets of points. There is no way of avoiding the introduction of constructional devices which seem to restrict the invariance of the theory, although once the theory is obtained it may be established in its full invariance.

Accordingly, we shall start our theory of randomness with a division of space, whether of one dimension or of more, into a denumerable assemblage of sub-sets. In one dimension, this division may be that into those intervals whose coördinates are terminating binary numbers, and in more dimensions, into those parallelepipeds with edges parallel to the axes and with terminating binary coördinates for the corner points. We then wish to find a self-consistent distribution-function for the mass in such a region, dependent only on the volume, and independent for non-overlapping regions.

This problem does not admit of a unique solution, although the solution becomes essentially unique if we adjoin suitable auxiliary conditions. Among these conditions, for example, is the hypothesis that the distribution is symmetric, as between positive and negative values, has a finite mean square, and that the measure of the set of α's for which $\mathfrak{F}(S;\alpha) > A$ is a continuous function of A.[5] Without going into such considerations, we shall assume directly that the measure of the set of instances in which the value of $\mathfrak{F}(S;\alpha)$ in a region S of measure M lies between a and $b > a$, is

$$\frac{1}{\sqrt{2\pi M}} \int_a^b \exp\left(-\frac{u^2}{2M}\right) du. \tag{57}$$

The formula

$$\frac{1}{\sqrt{2\pi M_1 M_2}} \int_{-\infty}^{\infty} \exp\left(-\frac{u^2}{2M_1} - \frac{(v-u)^2}{2M_2}\right) du = \frac{1}{\sqrt{M_1+M_2}} \exp\left(-\frac{v^2}{2(M_1+M_2)}\right) \tag{58}$$

shows the consistency of this assumption.

The distributions of mass among the sets of our denumerable assemblage may be mapped on the line segment $0 \leq \alpha \leq 1$, in such a way that the measure of the set of instances in which a certain contingency holds will go into a set of values of α of the same measure. This statement needs a certain amount of elucidation. To begin with, the only sets of instances whose measures we know are those determined by

$$\begin{aligned} a_1 &\leq \mathfrak{F}(S_1;\alpha) \leq b_1 \\ a_2 &\leq \mathfrak{F}(S_2;\alpha) \leq b_2 \\ &\cdots\cdots\cdots \\ a_n &\leq \mathfrak{F}(S_n;\alpha) \leq b_n; \end{aligned} \tag{59}$$

where $S_1, S_2, \cdots, S_n$ are to be found among our denumerable set of subdivisions of space. However, once we have established a correspondence between the measures of these specific sets of contingencies and their corresponding sets of values of α, we may use the measure of any measurable set of values of α to define the measure of its corresponding set of contingencies.

The correspondence between sets of contingencies and points on the line $(0, 1)$ is made by determining a hierarchy of sets of contingencies

$$\begin{aligned} a_1^{(m,n)} &\leq \mathfrak{F}(S_1^{(m,n)};\alpha) \leq b_1^{(m,n)} \\ &\cdots\cdots\cdots \\ a_\nu^{(m,n)} &\leq \mathfrak{F}(S_\nu^{(m,n)};\alpha) \leq b_\nu^{(m,n)}. \end{aligned} \tag{60}$$

[5] Cf. the recent investigations of Cramér and P. Lévy.

Let us call such a contingency $C_{m,n}$. If m is fixed, let all the contingencies $C_{m,n}$ $(n = 1, 2, \cdots)$ be mutually exclusive, and let them be finite in number. Let us be able to write

$$C_{m,n} = \sum_{k=1}^{N} C_{m+1,n_k}. \tag{61}$$

If S' is one of our denumerable sets of regions of space, let every $C_{m,n}$ with a sufficiently large index m be included in a class determined by a set of conditions concerning the mass on S' alone, and restricting it to a set of values lying in an interval (c, d), corresponding to an integral of form (57) and of arbitrarily small value. (Here d may be ∞, or c may be $-\infty$.) Let us put $C_{1,1}$ for the entire class of all possible contingencies, and let us represent it by the entire interval $(0, 1)$. Let us assume that $C_{m,n}$ has been mapped into an interval of length corresponding to its probability, in accordance with (57), and let this interval be divided in order of the sequence of their n_k's into intervals corresponding respectively to the component C_{m+1,n_k}'s, and of the same measure. Except for a set of points of zero measure, every point of the segment $(0, 1)$ of α will then be determined uniquely by the sequence of the intervals containing it and corresponding to the contingencies $C_{m,n}$ for successive values of m. This sequence will then determine uniquely (except in a set of cases corresponding to a set of values of α of zero measure) the value of $\mathfrak{F}(S_n; \alpha)$ for every one of our original denumerable set of sets S_n.

So far, everything that we have said has been independent of dimensionality. We now proceed to something belonging specifically to the one-dimensional case. If the original sets S_n are the sets of intervals with binary end-points, of such a form that they may be written in the binary scale

$$(d_1 d_2 \cdots d_k \cdot d_{k+1} \cdots d_l, d_1 \cdots d_k \cdot d_{k+1} \cdots d_l + 2^{-l-1}) \tag{62}$$

where $d_1, \cdots, d_l$ are digits which are either 0 or 1, then any interval whatever of length not exceeding $2^{-\mu}$ and lying in (a, b) (where a and b are integers) may be written as the sum of not more than two of the $(b-a)2^{\mu+2}$ intervals of form (62) lying in (a, b) and of length $2^{-\mu-1}$, not more than two of the $(b-a)2^{\mu+2}$ intervals of length $2^{-\mu-2}$, and so on. The probability that the value of $|\mathfrak{F}(S_n; \alpha)|$ should exceed A, or in other words, the measure of the set of α's for which it exceeds A, is

$$\frac{2}{\sqrt{2\pi m(S_n)}} \int_A^\infty \exp\left(-\frac{u^2}{2m(S_n)}\right) du = o\{\exp(-Am(S_n)^{-\frac{1}{2}})\}. \tag{63}$$

Now let us consider the total probability that the value of $|\mathfrak{F}(S_n; \alpha)|$ should exceed $2^{-(\mu+1)(\frac{1}{2}-\epsilon)}$ for any one of the $(b-a)2^{\mu+1}$ intervals of length

$2^{-\mu-1}$, or $2^{-(\mu+2)(\frac{1}{2}-\epsilon)}$ for any one of the $(b-a)2^{\mu+2}$ intervals of length $2^{-\mu-2}$, or so on. This probability can not exceed

$$(64) \qquad \sum_{k=1}^{\infty} (b-a)2^{\mu+k}\, o(\exp(-2^{(\mu+k)\epsilon})) = o(2^{\mu}\exp(-2^{(\mu+1)\epsilon})).$$

On the other hand, the sum of $|\mathfrak{F}(S_n;\alpha)|$ for all the $2+2+\cdots$ intervals must in any other case be equal to or less than

$$(65) \qquad 2\sum_{k=1}^{\infty} 2^{-(\mu+k)(\frac{1}{2}-\epsilon)} = O(2^{-\mu(\frac{1}{2}-\epsilon)}).$$

Thus there is a certain sense in which over a finite interval, and except for a set of values of α of arbitrarily small positive measure, the total mass in a sub-interval of length $\leq 2^{-\mu}$ tends uniformly to 0 with $2^{-\mu}$. On this basis, we may extend the functional $\mathfrak{F}(S;\alpha)$ to all intervals S. It is already defined for all intervals with terminating binary end-points. If (c,d) is any interval whatever, let $c_1, c_2, \cdots$ be a sequence of terminating binary numbers approaching c, and let $d_1, d_2, \cdots$ be a similar sequence approaching d. Then except for a fixed set of values of α of arbitrarily small measure,

$$(66) \qquad \lim_{m,n\to\infty} |\mathfrak{F}((c_n,d_n)+\alpha) - \mathfrak{F}((c_m,d_m);\alpha)| = 0,$$

and we may put

$$(67) \qquad \mathfrak{F}((c,d);\alpha) = \lim_{m\to\infty} \mathfrak{F}((c_m,d_m);\alpha).$$

Formula (57), and the fact that $\mathfrak{F}(S_1;\alpha)$ and $\mathfrak{F}(S_2;\alpha)$ vary independently for non-overlapping intervals S_1 and S_2, will be left untouched by this extension.

Thus if Σ is an interval and T^{λ} a translation through an amount λ, we can define $\mathfrak{F}(T^{\lambda}\Sigma;\alpha)$, and it will be equally continuous in λ over any finite range of λ except for a set of values of α of arbitrarily small measure. From this it follows at once that it is measurable in λ and α together. Furthermore, we shall have

$$(68) \qquad \begin{array}{l}\text{Measure of set of } \alpha\text{'s for which } \mathfrak{F}(T^{\lambda}\Sigma;\alpha) \text{ belongs to } C = \\ \text{Measure of set of } \alpha\text{'s for which } \mathfrak{F}(\Sigma;\alpha) \text{ belongs to } C.\end{array}$$

Thus $\mathfrak{F}(\Sigma;\alpha)$ is a homogeneous chaos. We shall call it *the pure chaos*.

If $\Phi\{\mathfrak{F}(\Sigma;\alpha)\}$ is a functional dependent on the values of $\mathfrak{F}(\Sigma;\alpha)$ for a finite number of intervals Σ_n, then if λ is so great that none of the intervals Σ_n overlaps any translated interval $T^{\lambda}\Sigma_m$, $\Phi\{\mathfrak{F}(\Sigma;\alpha)\}$ will have a distribution entirely independent of $\Phi\{\mathfrak{F}(T^{\lambda}\Sigma;\alpha)\}$. As every measurable functional may be approached in the L sense by such a functional, we see at once that $\mathfrak{F}(\Sigma;\alpha)$ is metrically transitive.[6]

[6] Except that the method of treatment has been adapted to the needs of § 7, the

7. Pure multidimensional chaos. In order to avoid notational complexity, we shall not treat the general multidimensional case explicitly, but shall treat the two-dimensional case by a method which will go over directly to the most general multidimensional case. If our initial sets S_n are the rectangles with terminating binary coördinates for their corners and sides parallel to the axes, and we replace (a, b) by the square with opposite vertices (p, q) and $(p + r,\ q + r)$, an argument of exactly the same sort as that which we have used in the last paragraph will show that except in a set of cases of total probability not exceeding

$$(69) \qquad \text{const.} \sum_{k=1}^{\infty} \sum_{l=1}^{\infty} 2^{\mu+k} 2^{l}\, o\left(\exp\left(-2^{(\mu+k+l)\epsilon}\right)\right) = o\left(2^{\mu} \exp\left(-2^{(\mu+1)\epsilon}\right)\right)$$

the sum of $|\mathfrak{F}(S_n; \alpha)|$ for a denumerable set of binary rectangles with base $\leq 2^{-\mu}$, of the form (62), and adding up to make a vertical interval lying in the square (p, q), $(p + r, q + r)$, must be equal to or less than

$$(65) \qquad 2 \sum_{k=1}^{\infty} 2^{-(\mu+k)\left(\frac{1}{2}-\epsilon\right)} = O\left(2^{-\mu\left(\frac{1}{2}-\epsilon\right)}\right).$$

If we now add this expression up for all the base intervals of type (62) necessary to exhaust a horizontal interval of magnitude not exceeding $2^{-\mu}$, we shall again obtain an expression of the form (65). It hence follows that if we take the total mass on the coördinate rectangles within a given square, this will tend to zero uniformly with their area, except for a set of values of α of arbitrarily small measure. From this point the two-dimensional argument, and indeed the general multidimensional argument, follows exactly the same lines as the one-dimensional argument. It is only necessary to note that if

$$a_n \to a,\ b_n \to b,\ c_n \to c,\ d_n \to d$$

then the rectangles (a_m, b_m), (c_m, d_m) and (a_n, b_n), (c_n, d_n) differ at most by four rectangles of small area.[7]

From this point on, we shall write $\mathfrak{P}(S; \alpha)$ for a pure chaos, whether in one or in more dimensions.

8. Phase averages in a pure chaos. If $f(P)$ is a measurable step-function, the definition of

$$(70) \qquad \int f(P)\, d_P \mathfrak{P}(S; \alpha)$$

---

results of this section have previously been demonstrated by the author. (*Proceedings of the London Mathematical Society*, 2, vol. 22 (1924), pp. 454-467).

[7] Here we represent a rectangle by giving two opposite corners.

is obvious, for it reduces to the finite sum

$$\sum_1^N f_n \mathcal{P}(S_n; \alpha) \tag{71}$$

where f_n are the N values assumed by $f(P)$, and S_n respectively are the sets over which these values are assumed. Let us notice that

$$\begin{aligned}(72)\quad \int_0^1 d\alpha \Big| \int f(P)\, d_P\mathcal{P}(S;\alpha)\Big|^2 &= \sum_{m=1}^N \sum_{n=1}^N \int_0^1 d\alpha f_m \bar{f}_n \mathcal{P}(S_m;\alpha)\mathcal{P}(S_n;\alpha)\\ &= \sum_1^N |f_n|^2 \int_0^1 (\mathcal{P}(S_n;\alpha))^2 d\alpha\\ &= \sum_1^N |f_n|^2 \frac{1}{\sqrt{2\pi m(S_n)}} \int_{-\infty}^{\infty} u^2 \exp\left(-\frac{u^2}{2m(S_n)}\right) du\\ &= \sum_1^N |f_n|^2 m(S_n) \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} u^2 e^{-(u^2/2)} du\\ &= \sum_1^N |f_n|^2 m(S_n) = \int |f(P)|^2 dV_P,\end{aligned}$$

the integral being taken over the whole of space. In other words, the transformation from $f(P)$ as a function of P, to $\int f(P)\, d_P\mathcal{P}(S;\alpha)$ as a function of α, retains distance in Hilbert space.[8] Such a transformation, by virtue of the Riesz-Fischer theorem, may always be extended by making limits in the mean correspond to limits in the mean. Thus both in the one-dimensional and in the many-dimensional case, we may *define*

$$\int f(P)\, d_P\mathcal{P}(S;\alpha) = \text{l. i. m.}_{n\to\infty} \int f_n(P)\, d_P\mathcal{P}(S;\alpha) \tag{73}$$

where $f(P)$ is a function belonging to L^2, and the sequence $f_1(P), f_2(P), \cdots$ is a sequence of step-functions converging in the mean to $f(P)$ over the whole of space. The definition will be unambiguous, except for a set of values of α of zero measure.

If S is any measurable set, we have

$$\begin{aligned}(74)\quad \int_0^1 d\alpha \{\mathcal{P}(S;\alpha)\}^n &= \frac{1}{\sqrt{2\pi m(S)}} \int_{-\infty}^{\infty} u^n \exp\left(-\frac{u^2}{2m(S)}\right) du\\ &= (m(S))^{n/2} \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} u^n e^{-(u^2/2)} du\\ &\begin{cases} = 0 \text{ if } n \text{ is odd}\\ = (m(S))^{n/2}(n-1)(n-3)\cdots 1 \text{ if } n \text{ is even.}\end{cases}\end{aligned}$$

[8] Cf. Paley, Wiener, and Zygmund, *Mathematische Zeitschrift*, vol. 37 (1933), pp. 647-668.

This represents $(m(S))^{n/2}$, multiplied by the number of distinct ways of representing n objects as a set of pairs. Remembering that if $S_1, S_2, \cdots, S_{2n}$ are non-overlapping, their distributions are independent, we see that if the sets $\Sigma_1, \Sigma_2, \cdots, \Sigma_{2n}$ are either totally non-overlapping, or else such that when two overlap, they coincide, we have [9]

$$(75) \quad \int_0^1 \mathcal{P}(\Sigma_1; \alpha) \cdots \mathcal{P}(\Sigma_n; \alpha) d\alpha = \Sigma\Pi \int_0^1 \mathcal{P}(\Sigma_j; \alpha) \mathcal{P}(\Sigma_k; \alpha) d\alpha,$$

where the product sign indicates that the $2n$ terms are divided into n sets of pairs, j and k, and that these factors are multiplied together, while the addition is over all the partitions of $1, \cdots, 2n$ into pairs. If $2n$ is replaced by $2n+1$, the integral in (75) of course vanishes.

Since $\mathcal{P}(S; \alpha)$ is a linear functional of sets of points, and since both sides of (75) are linear with respect to each $\mathcal{P}(\Sigma_k; \alpha)$ separately, (75) still holds when $\Sigma_1, \Sigma_2, \cdots, \Sigma_{2n}$ can be reduced to sums of sets which either coincide or do not overlap, and hence holds for all measurable sets.

Now let $f(P_1, \cdots, P_n)$ be a measurable step-function: that is, a function taking only a finite set of finite values, each over a set of values $P_1, \cdots, P_n$ which is a product-set of measurable sets in each variable P_k. Clearly we may define

$$(76) \quad \int \cdots \int f(P_1, \cdots, P_n) d_{P_1}\mathcal{P}(S; \alpha) \cdots d_{P_n}\mathcal{P}(S; \alpha)$$

in a way quite analogous to that in which we have defined (70), and we shall have

$$(77) \quad \int_0^1 d\alpha \int \cdots \int f(P_1, \cdots, P_n) d_{P_1}\mathcal{P}(S; \alpha) \cdots d_{P_n}\mathcal{P}(S; \alpha) = \Sigma \int \cdots \int f(P_1, P_1, P_2, P_2, \cdots, P_n, P_n) dV_{P_1} \cdots dV_{P_n}$$

where the summation is carried out for all possible divisions of the $2n$ P's into pairs. Similarly in the odd case

$$(78) \quad \int_0^1 d\alpha \int \cdots \int f(p_1, \cdots, P_{2n+1}) d_{P_1}\mathcal{P}(S; \alpha) \cdots d_{P_n}\mathcal{P}(S; \alpha) = 0.$$

We may apply (77) to give a meaning to

$$(79) \quad \int_0^1 d\alpha \left| \int \cdots \int f(P_1, \cdots, P_n) d_{P_1}\mathcal{P}(S; \alpha) \cdots d_{P_n}\mathcal{P}(S; \alpha) \right|^2.$$

If $f(P_1, \cdots, P_n)$ is a measurable step-function, and

[9] Cf. Paley, Wiener, and Zygmund, *loc. cit.*, formula (2.05).

$$(80)\qquad |f(P_1,\cdots,P_n)| \leq |f_1(P_1)\cdots f_n(P_n)|\ ; \qquad \int |f_k(P)|^2 dV_P \leq A \qquad (k=1,2,\cdots)$$

we shall have

$$(81)\qquad \int_0^1 d\alpha \left| \int\cdots\int f(P_1,\cdots P_n)\, d_{P_1}\mathfrak{P}(S;\alpha)\cdots d_{P_n}\mathfrak{P}(S;\alpha)\right|^2 \leq A^n(2n-1)(2n-3)\cdots 1.$$

If now $f(P_1,\cdots,P_n)$ is an integrable function satisfying (80), but not necessarily a step-function, let

$$(82)\quad f(\nu;P_1,\cdots,P_n)=\frac{1}{\nu}\operatorname{sgn} f(P_1,\cdots,P_n)\,[\nu f(P_1,\cdots,P_n)\operatorname{sgn} f(P_1,\cdots,P_n)].$$

Clearly almost everywhere

$$(83)\qquad \int\cdots\int f(\nu;P_1,\cdots,P_n)\, d_{P_1}\mathfrak{P}(S;\alpha)\cdots d_{P_n}\mathfrak{P}(S;\alpha) \leq \prod_1^n \int f_k(P)\, d_P\mathfrak{P}(S;\alpha)$$

and

$$(84)\qquad \overline{\lim_{\mu,\nu\to\infty}} \left| \int\cdots\int f(\mu;P_1,\cdots,P_n)\, d_{P_1}\mathfrak{P}(S;\alpha)\cdots d_{P_n}\mathfrak{P}(S;\alpha) - \int\cdots\int f(\nu;P_1,\cdots,P_n)\, d_{P_1}\mathfrak{P}(S;\alpha)\, d_{P_2}\mathfrak{P}(S;\alpha)\right| \leq \prod_1^n \int f_k(P)\, d_P\mathfrak{P}(S;\alpha)\left[\epsilon+\sum_1^n\left\{\frac{\int_R f_k(P)\, d_P\mathfrak{P}(S;\alpha)}{\int f_k(P)\, d_P\mathfrak{P}(S;\alpha)}\right\}\right]$$

where R represents the exterior of a sphere of arbitrarily large volume. Let it be noted that both the numerator and the denominator of this fraction have Gaussian distributions, but that the mean square value of the numerator is arbitrarily small. Thus except for a set of values of α of arbitrarily small measure, the right side of expression (84) is arbitrarily small, so that we may write

$$(85)\qquad \lim_{\mu,\nu\to\infty}\left\{\int\cdots\int f(\mu;P_1,\cdots,P_n)\, d_{P_1}\mathfrak{P}(S;\alpha)\cdots d_{P_n}\mathfrak{P}(S;\alpha) - \int\cdots\int f(\nu;P_1,\cdots,P_n)\, d_{P_1}\mathfrak{P}(S;\alpha)\cdots d_{P_n}\mathfrak{P}(S;\alpha)\right\}=0.$$

Thus by dominated convergence,

$$(86)\qquad \lim_{\mu\to\infty}\int\cdots\int f(\mu;P_1,\cdots,P_n)\, d_{P_1}\mathfrak{P}(S;\alpha)\cdots d_{P_n}\mathfrak{P}(S;\alpha)$$

exists for almost all values of α, and we may write it by definition

$$\int \cdots \int f(P_1, \cdots, P_n) d_{P_1}\mathcal{P}(S;\alpha) \cdots d_{P_n}\mathcal{P}(S;\alpha). \tag{87}$$

This will clearly be unique, except for a set of values of α of zero measure. There will then be no difficulty in checking (77), (78), and (81).

9. Forms of chaos derivable from a pure chaos. Let us assume that $f(P)$ belongs to L^2, or that $f(P_1, \cdots, P_n)$ is a measurable function satisfying (80). Let us write $\widehat{PQ}$ for the vector in n-space connecting the points P and Q. Then the function

$$\int \cdots \int f(\widehat{PP_1}, \cdots, \widehat{PP_n}) d_{P_1}\mathcal{P}(S;\alpha) \cdots d_{P_n}\mathcal{P}(S;\alpha) = F(P;\alpha) \tag{88}$$

is a metrically transitive differentiable chaos. This results from the fact that $\mathcal{P}(S;\alpha)$ is a metrically transitive chaos, and that a translation of P generates a similar translation of all the points P_k. The sum of a finite number of functions of the type (88) is also a metrically transitive differentiable chaos. To show that $F(P;\alpha)$ is measurable in P and α simultaneously, we merely repeat the argument of (83)–(86) with both P and α as variables.

We shall call a chaos such as (88) a *polynomial chaos homogeneously of the n-th degree,* and a sum of such chaoses a *polynomial chaos* of the degree of its highest term. In this connection, we shall treat a constant as a chaos homogeneously of degree zero.

By the multidimensional ergodic theorem, if Φ is a functional such that

$$\int_0^1 |\Phi(F(P;\alpha))| \log^+ |\Phi(F(P;\alpha))| \, d\alpha < \infty \tag{89}$$

we shall have

$$\lim_{r\to\infty} \frac{1}{V(r)} \int_R \Phi(F(P;\alpha)) dV_P = \int_0^1 \Phi(F(P;\alpha)) d\alpha \tag{90}$$

for almost all values of α. Since the distribution of $F(P;\alpha)$ is dominated by the product of a finite number of independent Gaussian distributions, we even have

$$\int_0^1 |F(P;\alpha)|^n d\alpha < \infty \tag{91}$$

for all positive integral values of n. In a wide class of cases this enables us to establish a relation of the type of (89).

In formula (90), we have an algorithm for the computation of the right-hand side. For example, if

$$(92) \qquad F(P;\alpha) = \int f(\widehat{PP_1})\, d_{P_1}\mathfrak{P}(S;\alpha),$$

and $P+Q$ is the vector sum of P and Q, we have for almost all α,

$$(93) \quad \lim_{r\to\infty} \frac{1}{V(r)} \int_R F(P+Q;\alpha)\bar{F}(Q;\alpha)\, dV_Q = \int f(P+Q)\bar{f}(Q)\, dV_Q,$$

the integral being taken over the whole of space; if

$$(94) \qquad F(P;\alpha) = \int \cdots \int f(\widehat{PP_1}, \widehat{PP_2})\, d_{P_1}\mathfrak{P}(S;\alpha)\, d_{P_2}\mathfrak{P}(S;\alpha),$$

we have almost always

$$(95) \quad \lim_{r\to\infty} \frac{1}{V(r)} \int_R F(P+Q;\alpha)\bar{F}(Q;\alpha)\, dV_Q$$
$$= \left| \int f(Q,Q)\, dV_Q \right|^2 + \int\int f(P+Q, P+M)\bar{f}(Q,M)\, dV_Q dV_M$$
$$+ \int\int f(P+Q, P+M)\bar{f}(M,Q)\, dV_Q dV_M;$$

and if

$$(96) \quad F(P;\alpha) = \int\int\int f(\widehat{PP_1}, \widehat{PP_2}, \widehat{PP_3})\, d_{P_1}\mathfrak{P}(S;\alpha)\, d_{P_2}\mathfrak{P}(S;\alpha)\, d_{P_3}\mathfrak{P}(S;\alpha),$$

we have almost everywhere

$$(97) \quad \lim_{r\to\infty} \frac{1}{V(r)} \int_R F(P+Q;\alpha)\bar{F}(Q;\alpha)\, dV_Q$$
$$= \int\int\int \{ f(Q,Q,P+M)\bar{f}(M,S,S)$$
$$+ f(Q,Q,P+M)\bar{f}(S,M,S) + f(Q,Q,P+M)\bar{f}(S,S,M)$$
$$+ f(Q,P+M,Q)\bar{f}(M,S,S) + f(Q,P+M,Q)\bar{f}(S,M,S)$$
$$+ f(Q,P+M,Q)\bar{f}(S,S,M) + f(P+M,Q,Q)\bar{f}(M,S,S)$$
$$+ f(P+M,Q,Q)\bar{f}(S,M,S) + f(P+M,Q,Q)\bar{f}(S,S,M)$$
$$+ f(P+Q,P+M,P+S)\bar{f}(Q,M,S)$$
$$+ f(P+Q,P+M,P+S)\bar{f}(Q,S,M)$$
$$+ f(P+Q,P+M,P+S)\bar{f}(M,Q,S)$$
$$+ f(P+Q,P+M,P+S)\bar{f}(M,S,Q)$$
$$+ f(P+Q,P+M,P+S)\bar{f}(S,Q,M)$$
$$+ f(P+Q,P+M,P+S)\bar{f}(S,M,Q) \}\, dV_Q dV_M dV_S.$$

We have similar results in the non-homogeneous case. Thus if

$$(98) \qquad F(P;\alpha) = A + \int f(\widehat{PP_1})\, d_{P_1}\mathfrak{P}(S;\alpha)$$
$$+ \int\int g(\widehat{PP_1}, \widehat{PP_2})\, d_{P_1}\mathfrak{P}(S;\alpha)\, d_{P_2}\mathfrak{P}(S;\alpha),$$

we have almost everywhere

$$\lim_{r\to\infty}\frac{1}{V(r)}\int_R F(P+Q;\alpha)\bar{F}(Q;\alpha)dV_Q \tag{99}$$
$$= A\int \bar{g}(Q,Q)dV_Q + \bar{A}\int g(Q,Q)dV_Q$$
$$+ \int f(P+Q)\bar{f}(Q)dV_Q + \left|\int g(Q,Q)dV_Q\right|^2$$
$$+ \int\int g(P+Q,P+M)\bar{g}(Q,M)dV_Q dV_M$$
$$+ \int\int g(P+Q,P+M)\bar{g}(M,Q)dV_Q dV_M.$$

10. Chaos theory and spectra.[10] The function

$$\lim_{r\to\infty}\frac{1}{V(r)}\int_R F(P+Q)\bar{F}(Q)dV_Q = G(P) \tag{100}$$

occupies a central position in the theory of harmonic analysis. If it exists and is continuous for every value of P, the function $F(P)$ is said to have an n-dimensional spectrum. To define this spectrum, we put

$$F_r(P) = \begin{cases} F(P) \text{ on } R; \\ 0 \text{ elsewhere.} \end{cases} \tag{101}$$

It is then easy to show by an argument involving considerations like those of (49) that if

$$G_r(P) = \frac{1}{V(r)}\int_\infty F_r(P+Q)\bar{F}_r(Q)dV_Q, \tag{102}$$

the integral being taken over the whole of space, then we have

$$G(P) = \lim_{r\to\infty} G_r(P). \tag{103}$$

Since, if O is the point with zero coördinates, by the Schwarz inequality,

$$|G(P)| \leq G(O), \tag{104}$$

the limit in (103) is approached boundedly.

If now we put

$$\phi_r(U) = (2\pi)^{-(n/2)}V(r)^{-\frac{1}{2}}\,\text{l. i. m.}_{s\to\infty}\int_S F_r(P)e^{iU.P}dV_P \tag{105}$$

where S is the interior of a sphere of radius s about the origin, the n-fold Parseval theorem will give us

---

[10] Cf. N. Wiener, "Generalized harmonic analysis," *Acta Mathematica*, vol. 55 (1930).

$$(106) \qquad |\phi_r(U)|^2 = (2\pi)^{-n}\int_\infty G_r(P)e^{iU.P}dV_P.$$

If $M(U)$ is a function with an absolutely integrable Fourier transform, we shall have

$$(107) \qquad \int_{-\infty}^{\infty}|\phi_r(U)|^2 M(U)dV_U = (2\pi)^{-n}\int_\infty G_r(P)dV_P\int_\infty M(U)e^{iU.P}dV_U,$$

and hence

$$(108) \qquad \lim_{r\to\infty}\int_{-\infty}^{\infty}|\phi_r(U)|^2 M(U)dV_U = (2\pi)^{-n}\int G(P)dV_P\int_\infty M(U)e^{iU.P}dV_U$$

which will always exist. Let us put

$$(109) \qquad \mathfrak{M}\{M(U)\} = (2\pi)^{-n}\int_\infty G(P)dV_P\int_\infty M(U)e^{iU.P}dV_U.$$

If S is any set of points of finite measure, and $S(P)$ is its characteristic function, let us put

$$(110) \qquad \overline{\mathfrak{M}}(S) = \underset{M(U)\geq S(U)}{\text{l. u. b.}}\ \mathfrak{M}(M(U)),$$

and

$$(111) \qquad \underline{\mathfrak{M}}(S) = \underset{M(U)\leq S(U)}{\text{g. l. b.}}\ \mathfrak{M}(M(U)).$$

If $\underline{\mathfrak{M}}(S)$ and $\overline{\mathfrak{M}}(S)$ have the same value, we shall write it $\mathfrak{M}(S)$, and shall call it the *spectral mass* of F on S. It will be a non-negative additive set-function of S, and may be regarded as determining the spectrum of F.

If $f(P_1,\cdots,P_n)$ satisfies (80) and $F(P;\alpha)$ is defined as in (88), we know that for any given P,

$$(112) \qquad G(P;\alpha) = \lim_{r\to\infty}\frac{1}{V(r)}\int_R F(P+Q;\alpha)\bar{F}(Q;\alpha) = \int_0^1 F(P,\beta)\bar{F}(0,\beta)d\beta$$

for almost all values of α. This alone is not enough to assure that $F(P,\alpha)$ has a spectrum for almost all values of α, as the sum of a non-denumerable set of sets of zero measure is not necessarily of zero measure. On the other hand, except for a set of values of α of zero measure, $G(P,\alpha)$ exists for all points P with rational coördinates.

We may even extend this result, and assert that if

$$(113) \qquad F_\theta(P;\alpha) = \frac{1}{V(\theta)}\int_{\text{length of } \widehat{PS}\leq\theta} F(S;\alpha)dV_S$$

and

$$(114) \qquad G_\theta(P;\alpha) = \lim_{r\to\infty}\frac{1}{V(r)}\int_R F_\theta(P+Q;\alpha)\bar{F}_\theta(Q;\alpha)dV_Q,$$

then except for a set of values of α of zero measure, $G_\theta(P;\alpha)$ exists for all points P with rational coördinates and all rational parameters θ, and it is easily proved that for almost all values of α, as θ tends to 0 through rational values,

$$\lim_{\theta\to 0}\lim_{r\to\infty}\frac{1}{V(r)}\int_R |F_\theta(Q;\alpha)-F(Q;\alpha)|^2\,dV_Q=0. \tag{115}$$

Now, by the Schwarz inequality,

$$\begin{aligned}(116)\quad &\Big|\frac{1}{V(r)}\int_R F_\theta(P+Q;\alpha)\bar F_\theta(Q;\alpha)\,dV_Q\\ &-\frac{1}{V(r)}\int_R F_\theta(P_1+Q;\alpha)\bar F_\theta(Q;\alpha)\,dV_Q\Big|\\ &\le\Big\{G_\theta(O;\alpha)\Big(\frac{1}{V(r)}\int_R |F_\theta(P+Q;\alpha)-F_\theta(P_1+Q;\alpha)|^2\,dV_Q\Big)\Big\}^{\frac12}\\ &\le\Big\{G_\theta(O;\alpha)\Big(\frac{1}{V(r)}\int_R dV_Q\Big(\frac{1}{V(\theta)}\Big[\int_{|P+\widehat{Q,S}|\le\theta}-\int_{|P+\widehat{Q,S}|\le\theta}\Big]|F(S;\alpha|^2\,dV_S\Big)\\ &\quad\times\Big(\frac{1}{V(\theta)}\Big[\int_{|P+\widehat{Q,S}|\le\theta}-\int_{|P_1+\widehat{Q,S}|\le\theta}\Big]\,dV_S\Big)\Big\}^{\frac12}\\ &\quad\le G_\theta(O;\alpha)O(\widehat{|PP_1|^{\frac12}}).\end{aligned}$$

It thus follows that if (114) exists for a given θ and all P's with rational coördinates, it exists for that θ and all real P's whatever. We may readily show that

$$G_\theta(O;\alpha)\le G(O;\alpha). \tag{117}$$

By another use of the Schwarz inequality,

$$\begin{aligned}(118)\quad &\Big|\frac{1}{V(r)}\int_R F_\theta(P+Q;\alpha)\bar F_\theta(Q;\alpha)\,dV_Q\\ &\qquad-\frac{1}{V(r)}\int_R F(P+Q;\alpha)\bar F(Q;\alpha)\,dV_Q\Big|\\ &\le\frac{1}{V(r)}\int_R |F_\theta(P+Q;\alpha)-F(P+Q;\alpha)|\;|F_\theta(Q;\alpha)|\,dV_Q\\ &\quad+\frac{1}{V(r)}\int_R |F(P+Q;\alpha)|\;|F_\theta(Q;\alpha)-F(Q;\alpha)|\,dV_Q\\ &\le\{G(O;\alpha)\frac{1}{V(r)}\int_R |F_\theta(P+Q;\alpha)-F(P+Q;\alpha)|^2\,dV_Q\}^{\frac12}\\ &\quad+\{G(O;\alpha)\frac{1}{V(r)}\int_R |F_\theta(Q;\alpha)-F(Q;\alpha)|^2\,dV_Q\}^{\frac12}.\end{aligned}$$

Combining (115) and (118), we see that except for a set of values of α of zero measure, we have for all P,

$$G(P;\alpha)=\lim_{\theta\to 0}G_\theta(P;\alpha). \tag{119}$$

We thus have an adequate basis for spectrum theory. This will extend, not merely to functions $F(P, \alpha)$ defined as in (88), but to finite sums of such functions. It will even extend to the case of any differentiable chaos $F(P; \alpha)$, for which $F(P+Q; \alpha)\bar{F}(Q; \alpha)$ is an integrable function of α, and for which (115) holds. For a metrically transitive chaos, this latter will be true if

$$\lim_{\theta \to 0} \int_0^1 | F_\theta(P; \alpha) - F(P; \alpha) |^2 \, d\alpha = 0. \tag{120}$$

Under this assumption, we have proved that $F(P)$ has a spectrum, and the same spectrum, for all values of α.

This enables us to answer a question which has been put several times, as to whether there is any relation between the spectrum of a chaos and the distribution of its values. There is no unique relation of the sort. The function

$$\int g(P+Q)\bar{g}(Q) dV_Q, \tag{121}$$

where g belongs to L^2, may be so chosen as to represent any Fourier transform of a positive function of L, and if $f(P, Q, M)$ is a bounded step-function, the right-hand side of (97) will clearly be the Fourier transform of a positive function of L. In particular, let $f(P_1, P_2, P_3) = f(P_1)f(P_2)f(P_3)$,

$$F_1(P; \alpha) = \int f_1(\widehat{PP_1}) \, d_{P_1} \mathfrak{P}(S; \alpha) \tag{122}$$

and choose $f_1(Q)$ in such a way that

$$\int f_1(P+Q)\bar{f}_1(Q) dV_Q = \text{right-hand side of (97).} \tag{123}$$

Then

$$\int_0^1 (F_1(P; \alpha))^{2n} d\alpha = \left(\int_\infty | f_1(P) |^2 \, dV_P \right)^n (2n-1)(2n-3) \cdots 1 \tag{124}$$

and if $F(P, \alpha)$ is defined as in (96),

$$\int_0^1 (F(P; \alpha))^{2n} d\alpha = \left(\int | f(P) |^2 dV_P \right)^{3n} (6n-1)(6n-3) \cdots 1) \tag{125}$$

so that for all but at most one value of n,

$$\int_0^1 (F_1(P; \alpha)^{2n} d\alpha \neq \int_0^1 (F(P; \alpha))^{2n} d\alpha \tag{126}$$

and we obtain in F and F_1 two chaoses with identical spectra but different distribution functions. On the other hand, if

$$\int_\infty | f_1(P) |^2 \, dV_P = \int_\infty | f_2(P) |^2 \, dV_P, \tag{127}$$

the chaoses

$$\int f_1(\widehat{PP_1})\, d_{P_1}\mathfrak{P}(S;\alpha) \tag{128}$$

and

$$\int f_2(\widehat{PP_1})\, d_{P_1}\mathfrak{P}(S;\alpha) \tag{129}$$

will have the same distribution functions, but may have very different spectra.

11. The discrete chaos.[11] Let us now divide the whole of Euclidean n-space dichotomously into sets $S_{m,n}$, such that every two sets S_{m_1,n_1} and S_{m_2,n_2} have the same measure, and that each $S_{m,n}$ is made up of exactly two non-overlapping sets $S_{m+1,k}$. Let us divide all these sets into two categories, "occupied," and "empty." Let us require that the probability that a set be empty depend only on its measure, and that the probability that two non-overlapping sets be empty be the product of the probabilities that each be empty. Let us assume that both empty and occupied sets exist. Let every set contained in an empty set be empty, while if a set be occupied, let at least one-half always be occupied. We thus get an infinite class of schedules of emptiness and occupiedness, and methods analogous to those of paragraph 6 may be used to map the class of these schedules in an almost everywhere one-one way on the line $(0, 1)$ of the variable α, in such a way that the set of schedules for which a given finite number of regions are empty or occupied will have a probability equal to the measure of the corresponding set of values of α.

By the independence assumption, the probability that a given set $S_{m,n}$ be empty must be of the form $e^{-Am(S_{m,n})}$. If $S_{m,n}$ is divided into the 2^ν intervals $S_{m+\nu,n_1}, \cdots, S_{m+\nu,n_{2^\nu}}$ at the ν-th stage of sub-division, the probability that just one is occupied and the rest are empty is

$$2^\nu(1-\exp(-Am(S_{m,n})/2^\nu)\exp\left(-\frac{2^\nu-1}{2^\nu}Am(S_{m,n})\right). \tag{130}$$

This contingency at the $\nu+1$-st stage is a sub-case of this contingency at the ν-th stage. If we interpret probability to mean the same thing as the measure of the corresponding set of α's, then by monotone convergence, the probability that at every stage, all but one of the subdivisions of $S_{m,n}$ are empty, while the remaining one is occupied, will be the limit of (130), or

$$Am(S_{m,n})\exp(-Am(S_{m,n})). \tag{131}$$

[11] The ideas of this paragraph are related to discussions the author has had with Profes or von Neumann, and the main theorem is equivalent to one enunciated by the latter.

Such a series of stages of subdivision will have as its occupied regions exactly those which contain a given point.

The probability that the occupied regions are exactly those which contain two points is the probability that each half of $S_{m,n}$ contain exactly one point, plus the probability that one-half is empty, and that in the occupied half, each quarter will contain exactly one point, plus and so on. This will be

$$(132)\quad \left\{ \frac{Am(S_{m,n})}{2} \exp\left(-\frac{Am(S_{m,n})}{2}\right) \right\}^2 + 2\exp\left(-\frac{Am(S_{m,n})}{2}\right) \left\{ \frac{Am(S_{m,n})}{2} \exp\left(-\frac{Am(S_{m,n})}{4}\right)^2 + \cdots = \exp(-Am(S_{m,n})m(S_{m,n}))^2(\tfrac{1}{4}+\tfrac{1}{8}+\cdots) = \frac{(Am(S_{m,n}))^2}{2}\exp(-Am(S_{m,n})).$$

If the probability that the occupied regions are exactly those containing $k-1$ points is

$$\frac{(Am(S_{m,n}))^{k-1}}{(k-1)!}\exp(-Am(S_{m,n}))$$

then a similar argument will show that the probability that the occupied regions are exactly those containing k points will be

$$(133)\quad \sum_{j=1}^{k-1}\frac{1}{j!}\frac{1}{(k-j)!}(Am(S_{m,n}))^k\exp(-Am(S_{m,n}))\left(1+\frac{1}{2^{k-1}}+\frac{1}{4^{k-1}}+\cdots\right) = \frac{1}{k!}(2^k-2)\left(\frac{1}{1-\frac{1}{2^{k-1}}}\right)(Am(S_{m,n}))^k\exp(-Am(S_{m,n})) = \frac{1}{k!}(Am(S_{m,n}))^k\exp(-Am(S_{m,n})).$$

Thus by mathematical induction, the probability that the occupied regions are exactly those containing k points will be

$$\frac{1}{k!}(Am(S_{m,n}))^k\exp(-Am(S_{m,n}))$$

and the sum of this for all values of k will be

$$(134)\qquad \sum_0^\infty \frac{1}{k!}(Am(S_{m,n}))^k\exp(-Am(S_{m,n})) = 1.$$

In other words, except for a set of contingencies of probability zero, the occupied regions will be exactly those containing a given finite number of points.

We may proceed at once from the fact that the probability that a set S_1 contains exactly k points is

$$\frac{1}{k!}(Am(S_1))^k e^{-Am(S_1)}$$

while the probability that the non-overlapping set S_2 contains exactly k points is

$$\frac{1}{k!}(Am(S_2))^k e^{-Am(S_2)}$$

to the fact that the probability that the set $S_1 + S_2$ contains exactly k points is

$$\sum_0^k \frac{1}{j!}\frac{1}{(k-j)!}(Am(S_1))^j(Am(S_2))^{k-j}e^{-Am(S_1+S_2)} = \frac{1}{k!}(Am(S_1+S_2))^k e^{-Am(S_1+S_2)}. \tag{135}$$

From this, by monotone convergence, it follows at once that the probability that any set S which is the sum of a denumerable set of our fundamental regions $S_{m,n}$ should contain exactly k points is

$$\frac{1}{k!}(Am(S))^k e^{-Am(S)}.$$

It is then easy to prove this for all measurable sets S.

We are now in a position to prove that the additive functional $\mathfrak{D}(S;\alpha)$, consisting in the number of points in the region S on the basis of the schedule corresponding to α, is a homogeneous metrically transitive chaos. The rôle which continuity filled in paragraphs 6 and 7, of allowing us to show that $\mathfrak{F}_{y_1,\ldots,y_n}(S;\alpha)$ was measurable in $y_1,\cdots,y_n$ and α, is now filled by the fact that the probability that any of the points in a region lie within a very small distance of the boundary, is for any Jordan region the probability that a small region be occupied, and is small. The metric transitivity of the chaos results as before from the independence of the distribution in non-overlapping regions.

The discrete or Poisson chaos which we have thus defined is the chaos of an infinite random shot pattern, or the chaos of the gas molecules in a perfect gas in statistical equilibrium according to the old Maxwell statistical mechanics. It also has important applications to the study of polycrystalline aggregates, and to similar physical problems.

Two important formulae are

$$\int_0^1 \mathfrak{D}(S;\alpha)\,d\alpha = e^{-Am(S)}\sum_1^\infty \frac{k}{k!}(Am(S))^k = Am(S), \tag{136}$$

and

$$(137)\quad \int_0^1 (\mathcal{D}(S;\alpha))^2 d\alpha = e^{-Am(S)} \sum_1^\infty \frac{k^2}{k!} (Am(S))^k = (Am(S))^2 + Am(S).$$

Let it be noted that if we define

$$(138)\quad \int f(P) d_P \mathcal{D}(S;\alpha)$$

for a measurable step-function $f(P)$ as in (70), by

$$(139)\quad \sum_1^N f_n \mathcal{D}(S_n;\alpha),$$

(72) is replaced by

$$\begin{aligned}(140)\quad &\int_0^1 d\alpha \,|\int f(P) d_P \mathcal{D}(S;\alpha) - A\int_\infty f(Q) dV_Q\,|^2 \\ &= \sum_{m=1}^N \sum_{n=1}^N \int_0^1 d\alpha\, f_m \bar{f}_n \mathcal{D}(S_m;\alpha)\mathcal{D}(S_n;\alpha) \\ &\quad - 2\mathcal{R}\{\bar{A}\sum_{m=1}^\infty \int_0^1 d\alpha\, f_m \mathcal{D}(S_m;\alpha)\int_\infty f(Q) dV_Q + |A\int_\infty f(Q) dV_Q\,|^2\} \\ &= \sum_{m=1}^N |f_m|^2 Am(S_m) \\ &= A\int_\infty |f(P)|^2 dV_P.\end{aligned}$$

Thus the transformation from $f(P)$ as a function of P, to

$$(141)\quad \int f(P) d_P \mathcal{D}(S;\alpha) - A\int_\infty f(Q) dV_Q$$

as a function of α, retains distance in Hilbert space, apart from a constant factor, and if $f(P)$ belongs to L and L^2 simultaneously, and $\{f_n(P)\}$ is a sequence of step-functions converging in the mean both in the L sense and in the L^2 sense to $f(P)$, we may *define*

$$\begin{aligned}(142)\quad &\int f(P) d_P \mathcal{D}(S;\alpha) = A\int_\infty f(Q) dV_Q \\ &\qquad + \underset{n\to\infty}{\text{l. i. m.}} \left(\int f_n(P) d_P \mathcal{D}(S;\alpha) - A\int_\infty f_n(Q) dV_Q\right).\end{aligned}$$

As in the case of (73), this definition is substantially unique. We may prove the analogue of (93) in exactly the same way as (93) itself, and shall obtain

$$\begin{aligned}(143)\quad &\lim_{r\to\infty} \frac{1}{V(r)} \int_R \left\{\int_\infty (f(P+\widehat{Q)M};\alpha) d_M \mathcal{D}(S;\alpha) - A\int_\infty f(M) dV_M\right\} \\ &\qquad \times \left\{\int_\infty \bar{f}(\widehat{QM};\alpha) d_M \mathcal{D}(S;\alpha) - A\int_\infty \bar{f}(M) dV_M\right\} dV_Q \\ &\qquad = A^2 \int_\infty f(P+Q)\bar{f}(Q) dV_Q.\end{aligned}$$

As we may see by appealing to the theory of spectra, one interpretation of this in the one-dimensional case is the following: *If a linear resonator be set into motion by a haphazard series of impulses forming a Poisson chaos, the effect, apart from that of a constant uniform stream of impulses, will have the same power spectrum as the energy spectrum of the response of the resonator to a single impulse.*

12. The weak approximation theorem for the polynomial chaos. We wish to show that the chaoses of paragraph 9 are in some sense everywhere dense in the class of all metrically transitive homogeneous chaoses. We shall show that if $\mathfrak{F}(S;\alpha)$ is any homogeneous chaos in n dimensions, there is a sequence $\mathfrak{F}_k(S;\alpha)$ of polynomial chaoses as defined in paragraph 9, such that if $S_1, \cdots, S_\nu$ is any finite assemblage of bounded measurable sets in n-space selected from among a denumerable set, and

$$\int_0^1 |\mathfrak{F}(S_\lambda;\alpha)|^\mu d\alpha < \infty \qquad (\lambda = 1, 2, \cdots, \nu) \tag{144}$$

is finite, then

$$\int_0^1 \mathfrak{F}(S_1;\alpha)\cdots\mathfrak{F}(S_\nu;\alpha)d\alpha = \lim_{n\to\infty}\int_0^1 \mathfrak{F}_n(S_1;\alpha)\cdots\mathfrak{F}_n(S_\nu;\alpha)d\alpha. \tag{145}$$

We first make use of the fact that if the probability that a quantity u be greater in absolute value than A, be less than

$$\frac{2}{\sqrt{2\pi B}}\int_A^\infty e^{-(u^2/2B)}du, \tag{146}$$

then if $\psi(u)$ is any even measurable function bounded over $(-\infty, \infty)$, we may find a polynomial $\psi_\epsilon(u)$, such that the mean value of

$$|\psi(u) - \psi_\epsilon(u)|^n, \tag{147}$$

which will be

$$\frac{1}{\sqrt{2\pi B}}\int_{-\infty}^\infty |\psi(u) - \psi_\epsilon(u)|^n e^{-(u^2/2B)}du, \tag{148}$$

is less than ϵ. Since it is well known that if $\phi(u)$ is a continuous function vanishing outside a finite interval, and

$$\sum_1^\infty A_n H_n(u)e^{-(u^2/2)} \tag{149}$$

is the series for $\phi(u)$ in Hermite functions, then we have uniformly

$$\phi(u) = \lim_{t \to 1-0} \sum_{1}^{\infty} A_n t^n H_n(u) e^{-(u^2/2)}, \tag{150}$$

to establish the existence of $\psi_\epsilon(u)$, we need only prove it in the case in which

$$\psi(u) = u^k e^{-Cu^2} \tag{151}$$

for an arbitrarily small value of C: as for example for $C = 1/4n_1B$. We shall then have

$$\left| \psi(u) - u^k \sum_{0}^{N} - \frac{(cu^2)^k}{k!} \right| \leq |u|^k \sum_{0}^{\infty} \frac{(cu^2)^k}{k!} = |u|^k e^{u^2/4n_1B}, \tag{152}$$

so that by dominated convergence, and if we take N large enough, we may make

$$\frac{1}{\sqrt{2\pi B}} \int_{-\infty}^{\infty} |\psi(u) - \psi_\epsilon(u)|^n e^{-(u^2/2B)} du < \epsilon \qquad (n \leq n_1). \tag{153}$$

Now let

$$\psi_K(u) = \begin{cases} 0 & (|u| < K); \\ 1 & (|u| \geq K); \end{cases} \tag{154}$$

and let us put

$$\mathfrak{G}(P; \alpha) = \frac{1}{V(r)} \mathfrak{P}(S; \alpha) \quad (S = \text{interior of } |\widehat{QP}| \leq r). \tag{155}$$

The chaos

$$\mathfrak{L}(P; \alpha) = \psi_K(\mathfrak{G}(P; \alpha)) \tag{156}$$

may then be approximated by polynomial chaoses in such a way as to approximate simultaneously to all polynomials in $\mathfrak{L}(P; \alpha)$ by corresponding polynomials in the approximating chaoses. Since the distribution of the values of $\mathfrak{G}(P; \alpha)$ will be Gaussian, with a root mean square value proportional to a power of r, and $\mathfrak{G}(P; \alpha)$ will be independent in spheres of radius η about two points P_1 and P_2 more remote from each other than $2r + 2\eta$, it follows that if we take K to be large enough, we may make the probability that $\mathfrak{L}(P; \alpha)$ differs from 0 between two spheres of radii respectively $r + \eta$ and H about a given point where it differs from 0, as small as we wish.

We now form the new chaos

$$\int_{|\widehat{PQ}| < x} \mathfrak{L}(Q; \alpha) dV_Q, \tag{157}$$

which we may also approximate, with all its polynomial functionals, by a sequence of polynomial chaoses. The use of polynomial approximations

tending boundedly to a step function over a finite range will show us that this is also true of the chaos determined by

$$\psi_\gamma\Big(\int_{\widehat{|PQ|}<x} \mathfrak{L}(Q;\alpha)\,dV_Q = \mathfrak{M}(P;\alpha). \tag{158}$$

By a proper choice of the parameters, this can be made to have arbitrarily nearly all its mass uniformly distributed over regions arbitrarily near to arbitrarily small spheres, all arbitrarily remote from one another, except in an arbitrarily small fraction of the cases. We then form

$$\frac{1}{(2\pi k)^{n/2}}\int_\infty (\mathfrak{M}(Q;\alpha)+\delta)\exp\left(-\frac{|\widehat{PQ}|^2}{2k}\right)dV_Q = \mathfrak{N}(P;\alpha) \tag{159}$$

where δ is taken to be very small. This chaos again, as far as all its polynomial functionals are concerned, will be approximable by polynomial chaoses. Since it is bounded away from 0 and ∞, and since over such a range the function $1/x$ may be approximated uniformly by polynomials, it follows that in our sense,

$$1/\mathfrak{N}(P;\alpha) \tag{160}$$

is approximable by polynomial chaoses.

If $\varpi(P)$ is any measurable function for which arbitrarily high moments are always finite, it is easy to show that

$$\frac{1}{(2\pi k)^{n/2}}\int_\infty \varpi(Q)\,(\mathfrak{M}(Q;\alpha)+\delta)\exp\left(-\frac{|\widehat{PQ}|^2}{2k}\right)dV_Q = \mathfrak{W}(P;\alpha) \tag{161}$$

is approximable by polynomial chaoses. Multiplying expressions (160) and (161), it follows that

$$\mathfrak{W}(P;\alpha)/\mathfrak{N}(P;\alpha) = \mathfrak{U}(P;\alpha) \tag{162}$$

is approximable by polynomial chaoses.

If A is a large enough constant, depending on the choice of the constant ϵ, we have

$$\begin{aligned}\frac{1}{(2\pi k)^{n/2}}\int_{|P|>A}\exp\left(-\frac{|P|^2}{2k}\right)dV_P &\\ = \int_A^\infty x^n e^{-(x^2/2k)}dx \Big/ \int_0^\infty x^n e^{-(x^2/2k)}dx &\\ < \frac{1}{(2\pi k)^{n/2}}\exp\left(-\frac{(A-\epsilon)^2}{2k}\right).&\end{aligned} \tag{163}$$

Thus by the proper choice of the parameters of $\mathfrak{M}(P;\alpha)$, if we take k small enough and then δ small enough, the chaos (162) will consist as nearly as we wish, from the distribution standpoint, of an infinite assemblage of convex cells of great minimum dimension, in each of which the function $\varpi(P)$ is repeated, with the origin moved to some point remote from the boundary.

Now let $\mathfrak{F}(S;\alpha)$ be a metrically transitive homogeneous chaos. Let us form

$$\mathfrak{F}(r;S;\alpha) = \frac{1}{V(r)} \int_R \mathfrak{F}_{x_1,\ldots,x_n}(S;\alpha)\, dx_1 \cdots dx_n. \tag{164}$$

Clearly by the fundamental theorem of the calculus, over any finite region in $(x_1,\cdots,x_n)$, we shall have for almost all points and almost all values of α,

$$\mathfrak{F}(S;\alpha) = \lim_{r\to 0} \mathfrak{F}(r;S;\alpha); \tag{165}$$

and if (144) holds, it is easy to show that

$$\int_0^1 |\mathfrak{F}(r;S;\alpha)|^n d\alpha < \text{const.} \tag{166}$$

From this it follows that

$$\lim_{r\to 0} \int_0^1 |\mathfrak{F}(r;S;\alpha) - \mathfrak{F}(S;\alpha)|^n d\alpha = 0 \tag{167}$$

and by the ergodic theorem, except for a set of values of α of zero measure, as r tends to 0 through a denumerable set of values,

$$\lim_{r\to 0} \frac{1}{V(r)} \int_R |\mathfrak{F}_{x_1,\ldots,x_n}(r;S;\alpha) - \mathfrak{F}_{x_1,\ldots,x_n}(S;\alpha)|^n dx_1 \cdots dx_n = 0. \tag{168}$$

With this result as an aid, enabling us to show that the distribution of $\mathfrak{F}(S;\alpha)$ is only slightly affected by averaging within a small sphere with a given radius, or even within any small region near enough to a small sphere with a given radius, we may proceed as in (161) and (162) and form the chaos

$$\mathfrak{F}_k(S;\alpha) = \frac{1}{\mathfrak{N}(\mathfrak{P};\alpha)(2\pi k)^{n/2}} \int_\infty \mathfrak{F}_{x_1,\ldots,x_n}(S;\beta) \times (\mathfrak{M}(x_1,\cdots,x_n;\alpha)+\delta) \exp\frac{(-\sum_1^n x_j^2)}{2k}\, dx_1 \cdots dx_n. \tag{169}$$

For almost all β, in each of the large cells of this chaos, (169) will have as nearly as we wish the same distribution as some $\mathfrak{F}_{x_1, \cdots, x_n}(S; \alpha)$, where $(x_1, \cdots, x_n)$ lies in the interior of the cell, remote from the boundary. These cells may so be determined that except for those filling an arbitrarily small proportion of space, all are convex regions with a minimum dimension greater than some given quantity.

To establish (143), it only remains to show that the average of a quantity depending on a chaos over a large cell tends to the same limit as its average over a large sphere. To show this, we only need to duplicate the argument of paragraph 4, where we prove the multidimensional ergodic theorem, for large pyramids with the origin as a corner, instead of for large spheres about the origin. We may take the shapes and orientations of these pyramids to form a denumerable assemblage, from which we may pick a finite assemblage which will allow us to approach as closely as we want to any cell for which the ratio of the maximum to the minimum distance from the origin within it does not exceed a given amount. It is possible to show that by discarding cells whose measure is an arbitrarily small fraction of the measure of all space, the remaining cells will have this property.

13. The physical problem. The transformation of a chaos. The statistical theory of a homogeneous medium, such as a gas or liquid, or a field of turbulence, deals with the problem, given the statistical configuration and velocity distribution of the medium at a given initial time, and the dynamical laws to which it is subject, to determine the configuration at any future time, with respect to its statistical parameters. This of course is not a problem in the first instance of the history of the individual system, but of the entire ensemble, although in proper cases it is possible to show that almost all systems of the ensemble do actually share the same history, as far as certain specified statistical parameters are concerned.

The dynamical transformations of a homogeneous system have the very important properties, that they are independent of any choice of origin in time or in space. Leaving the time variable out of it, for the moment, the simplest space transformations of a homogeneous chaos $\mathfrak{F}(S; \alpha)$ which have this property are the polynomial transformations which turn it into

$$
\begin{aligned}
&K_0 + \int K_1(x_1 - y_1, x_2 - y_2, \cdots, x_n - y_n)\mathfrak{F}_{y_1, \cdots, y_n}(S; \alpha)\, dy_1 \cdots dy_n \\
&\quad + \cdots \\
&\quad + \int \cdots \int K_\nu(x_1 - y_1^{(1)}, \cdots, x_n - y_n^{(1)}, \cdots, x_1 - y_1^{(\nu)}, \cdots, \\
&\qquad x_n - y_n^{(\nu)})\mathfrak{F}_{y_1^{(1)}, \cdots, y_n^{(1)}}(S; \alpha) \cdots \\
&\qquad\qquad \mathfrak{F}_{y_1^{(\nu)}, \cdots, y_n^{(\nu)}}(S; \alpha)\, dy_1^{(1)} \cdots dy_n^{(\nu)}.
\end{aligned} \tag{170}
$$

These are a sub-class of the general class of polynomial transformations

$$
\begin{aligned}
&K_0 + \int K_1(x_1, \cdots, x_n; y_1, \cdots, y_n)\mathfrak{F}_{y_1, \ldots, y_n}(S; \alpha)\, dy_1 \cdots dy_n \\
&\quad + \cdots \\
&\quad + \int \cdots \int K_\nu(x_1, \cdots, x_n; y_1^{(1)}, \cdots, y_n^{(1)}; \cdots; y_1^{(\nu)}, \cdots, y_n^{(\nu)}) \\
&\qquad \times \mathfrak{F}_{y_1^{(1)}, \ldots, y_n^{(1)}}(S; \alpha) \cdots \mathfrak{F}_{y_1^{(\nu)}, \ldots, y_n^{(\nu)}}(S; \alpha)\, dy_1^{(1)} \cdots dy_n^{(\nu)}.
\end{aligned} \tag{171}
$$

If a transformation of type (171) is invariant with respect to position in space, it must belong to class (170). On the other hand, in space of a finite number of dimensions and in any of the ordinary spaces of an infinite number of dimensions, polynomials are a closed set of functions, and hence every transformation may be approximated by a transformation of type (171).

A polynomial transformation such as (170) of a polynomial chaos yields a polynomial chaos. If then we can approximate to the state of a dynamical system at time 0 by a polynomial chaos, and approximate to the transformation which yields its status at time t by a polynomial transformation, we shall obtain for its state at time t, the approximation of another polynomial chaos. The theory of approximation developed in the last section will enable us to show this.

On the other hand, the transformation of a dynamical system induced by its own development is infinitely subdivisible in the time, and except in the case of linear transformations, this is not a property of polynomial transformations. Furthermore, when these transformations are non-linear, they are quite commonly not infinitely continuable in time. For example, let us consider the differential equation

$$\frac{\partial u}{\partial t} + u\frac{\partial u}{\partial x} = 0. \tag{172}$$

This corresponds to the history of a space-distribution of velocity transferred by particles moving with that velocity. Its solutions are determined by the equation

$$u(x, t) = u(x - tu(x, t), 0), \tag{173}$$

or if ψ is the inverse function of $u(x, 0)$,

$$x - tu(x, t) = \psi(u(x, t)). \tag{174}$$

Manifestly, if two particles with different velocities are allowed to move long

enough to allow their space-time paths to cross, $u(x, t)$ will cease to exist as a single-valued function. This will always be the case for *some* value of t if $u(x, 0)$ is not constant, and for almost all values of t and α if it is a polynomial chaos.

By Lagrange's formula, (174) may be inverted into

$$u(x, t) = \sum_0^\infty \frac{(-t)^n}{n!} \frac{\partial^n}{\partial x^n} \{(u(x, 0))^n\}. \tag{175}$$

In a somewhat generalized sense, the partial sums of this formally represent polynomial transformations of the initial conditions. However, it is only for a very special sort of bounded initial function, and for a finite value of the time, that they converge. It is only in this restricted sense that the polynomial transformation represents a true approximation to that given by the differential equation.

It will be seen that the useful application of the theory of chaos to the study of particular dynamical chaoses involves a very careful study of the existence theories of the particular problems. In many cases, such as that of turbulence, the demands of chaos theory go considerably beyond the best knowledge of the present day. The difficulty is often both mathematical and physical. The mathematical theory may lead inevitably to a catastrophe beyond which there is no continuation, either because it is not the adequate presentation of the physical facts; or because after the catastrophe the physical system continues to develop in a manner not adequately provided for in a mathematical formulation which is adequate up to the occurrence of the catastrophe; or lastly, because the catastrophe does really occur physically, and the system really has no subsequent history. The hydrodynamical investigations required in the case of turbulence are directly in the spirit of the work of Oseen and Leray, but must be carried much further.

The study of the history of a mechanical chaos will then proceed as follows: we first determine the transformation of the initial conditions generated by the dynamics of the ensemble. We then determine under what assumptions the initial conditions admit of this transformation for either a finite or an infinite interval of time. Then we approximate to the transformation for a given range of values of the time by a polynomial transformation. Then, having regard to a definition of distance between two functions determined by the transformation, we approximate to the initial chaos by a polynomial chaos. Next we apply the polynomial transformation to the polynomial chaos, and obtain an approximating polynomial chaos at time t. Finally, we apply our algorithm of the pure chaos to determine the averages

of the statistical parameters of this chaos, and express these as functions of the time.

The results of such an investigation belong to a little-studied branch of statistical mechanics: the statistical mechanics of systems not in equilibrium. To study the classical, equilibrium theory of statistical mechanics by the methods of chaos theory is not easy. As yet we lack a method of representing all forms of homogeneous chaos, which will tell us by inspection when two differ merely by an equimeasure transformation of the parameter of distribution. In certain cases, in which the equilibrium is stable, the study of the history of a system with an arbitrary initial chaos will yield us for large values of t an approximation to equilibrium, but this will often fail to be so, particularly in the case of differentiable chaoses, or the only equilibrium may be that in which the chaos reduces to a constant.

MASSACHUSETTS INSTITUTE OF TECHNOLOGY.

ON HILL'S PERIODIC LUNAR ORBIT.

By AUREL WINTNER.

Hill's desideratum. This paper deals with a question concerning the mathematical foundations of Hill's lunar theory, a theory developed in the first volume of this Journal [1] and considered by Poincaré as representing a turning point in the history of celestial mechanics.

From the mathematical point of view, the most original, but also the boldest, step in Hill's lunar theory consists of abandoning the classical method of variation of constants, i. e., Lagrange's theory of perturbations, and replacing it by successive applications of the method of infinitely many variables. Among mathematicians, Hill's name is remembered mainly for his theory of the lunar perigee, a theory depending on a linear Eigenwertproblem of specific type and leading to an infinite system of homogeneous linear equations. Hill's formal treatment of this problem by means of transcendental entire functions which are defined by infinite determinants has already been justified by Poincaré and has lead, via the investigations of Fredholm, to Hilbert's theory of linear transformations. Thus it is understandable that the introduction of the theory of infinite linear systems is generally considered as Hill's main contribution to pure analysis.

Actually, Hill's fundamental paper (*loc. cit.*[1]) does by no means deal with the linear Eigenwertproblem which leads to infinite determinants, but rather with a non-linear problem which leads to an infinite non-linear and non-recursive system of equations for infinitely many unknown Fourier constants; a system which seems to involve steps essentially more delicate than those which are implied by the application of infinite determinants in the linear problem. Furthermore, the linear problem presupposes that the non-linear problem has already been solved. In fact, the linear problem merely concerns the infinitesimal isoenergetic normal displacements of those periodic solutions of the non-linear equations of motion which lead to the infinite system of non-linear conditions.

---

[1] G. W. Hill, "Researches in the lunar theory," *American Journal of Mathematics*, vol. 1 (1878), pp. 5-26, 129-147, 245-260.

For a detailed presentation of Hill's programme in lunar theory, cf. pp. 195-236 (Chap. XI) of E. W. Brown's Treatise on the Lunar Theory (Cambridge, 1896), where further references may be found.

Hill was well aware of the novelty of the mathematical problem involved in his infinite non-linear implicit system of equations: "I regret that, on account of the difficulty of the subject . . . it does not appear that anything in the writings of Cauchy will help us to the conditions of convergence" (*loc. cit.*[1], p. 8). On the other hand, not only were Hill's numerical results in perfect agreement with the observed motion of the Moon, but also the terms of his expansions, obtained from the infinite non-linear system of equations in a formal and approximative manner, have shown, in contrast with the corresponding expansions of the other lunar theories, a surprisingly rapid decrease in absolute value; so that Hill (*loc. cit.*[1], p. 8) describes the probable situation as follows: "There cannot be a reasonable doubt that, in all cases, where we are compelled to employ infinite series in the solution of a problem, analysis is capable of being perfected to the point of showing us within what limits our solution is legitimate, and also giving us a limit which its error cannot surpass."

The analytical problem. The problem implied by this desideratum of Hill was treated by the writer several years ago.[2] The method consisted in proving (*loc. cit.*[2], pp. 263-265) a general existence theorem on infinite non-linear implicit systems which turned out to be the appropriate tool for dealing with Hill's problem and, at the same time, gave explicit estimates of the error terms which, in Hill's procedure, result by stopping the process of the comparison of coefficients at a certain stage.

The object of the present paper is to simplify this direct existence and convergence proof in such a way as to make it adaptable to sharper estimates. For this end, it will be convenient to use an existence theorem on infinite non-linear implicit systems which is somewhat more general[3] than the one used *loc. cit.*[2], and takes into account the peculiar situation in orders of magnitude which is an essential point in the problem. The proof of this appropriate existence theorem hardly differs from the proof of the more particular existence theorem which was used *loc. cit.*[2] and will, therefore, be omitted. Nevertheless, the paper is readable in itself.

Due to this general existence theorem, the treatment of Hill's problem becomes free of any ad hoc device, proceeds exactly along the lines which were followed by Hill in a formal way, and, in addition, takes into account the successive decrease of the orders of magnitude; thus proving, and also explaining, the convergence of Hill's process and supplying the realization of Hill's desideratum.

---

[2] A. Wintner, "Zur Hill'schen Theorie der Variation des Mondes," *Mathematische Zeitschrift*, vol. 24 (1925), pp. 259-265.

[3] Cf. A. Wintner, "Über die Differentialgleichungen der Himmelsmechanik," *Mathematische Annalen*, vol. 96 (1926), pp. 284-312.

It should be mentioned that the question as to the existence of Hill's family of periodic lunar orbits is implied by, but is not identical with, the question as to the legitimacy of Hill's bold method of infinitely many variables. In fact, the question as to the existence of these periodic orbits can be settled, according to Birkhoff, in a rather simple and intuitive manner.[4]

The equations of condition. Let x, y be the geocentric coördinates of the Moon in the rotating (i. e., synodic) Cartesian plane of Hill's restricted problem of three bodies. Then, if t denotes the time and m ($\gtrless 0$) an integration constant which varies in the neighborhood of $m = \pm 0$, Hill's aim (*loc. cit.*[2], pp. 129-131) is to find a continuous family of periodic solutions of the form

$$(1) \qquad x = \sum_{i=-\infty}^{+\infty} a_i \cos(2i+1)t/m, \qquad y = \sum_{i=-\infty}^{+\infty} a_i \sin(2i+1)t/m,$$

where the Fourier coefficients a_i are unknown and are to be determined as real analytic functions,

$$(2) \qquad a_i = a_i(m) \qquad (i = 0, \pm 1, \pm 2, \cdots)$$

of the integration constant m, which is the parameter of the family (1).

One can replace m by the varying period

$$(3) \qquad T = 2\pi m \qquad (m \gtrless 0)$$

of (1) and also by Jacobi's constant $C = C(m)$, i. e., by the (relative) energy, where, however, $C(m) \to +\infty$ whether $m \to +0$ or $m \to -0$, i. e., whether the solution belonging to a small m is direct or retrograde. The numerical value, m_0, of m in the case of the Moon is, of course, positive and somewhat less than $1/12$:

$$(4) \qquad m_0 = 0.08084 \cdots$$

(cf. *loc. cit.*[1], p. 146).

Now, Hill's method of calculating the unknown family (1) of periodic solutions, i. e., the infinitely many unknown functions (2), consists simply of a direct substitution of (1) into the differential equations of motion and their Jacobi integral. This gives, by comparison of the coefficients in the resulting Fourier series, infinitely many implicit conditions for the infinitely many functions (2). Since the equations of motion are non-linear, so are these implicit conditions. In fact, Hill finds after several ingenious transformations

---

[4] G. D. Birkhoff, "The restricted problem of three bodies," *Rendiconti del Circolo Matematico di Palermo*, vol. 39 (1915), Part III, § 12. Cf. also A. Wintner, Sortengenealogie, Hekubakomplex und Gruppenfortsetzung," *Mathematische Zeitschrift*, vol. 34 (1931), § 5 (Fig. 1) and § 7. A Russian paper by A. Liapounoff is reviewed in the *Fortschritte der Mathematik*, vol. 26 (1895), p. 1103.

that the infinite system of conditions which are to be satisfied by the functions (2) consists, on the one hand, of the equation [5]

$$(5) \qquad \{\sum_{i=-\infty}^{+\infty} a_i\}^2 \sum_{i=-\infty}^{+\infty} \{4i^2 + 4i + 1 + \overline{4i + \tfrac{1}{2}m} + 3m^2\}a_i = m^2$$

and, on the other hand, of the simultaneous infinite system [6]

$$(6) \qquad \sum_{i=-\infty}^{+\infty} \{[j,i]a_i a_{i-j} + m^2[j]a_i a_{-i+j-1} + m^2(j)a_i a_{-i-j-1}\} = 0; \qquad (j = \pm 1, \pm 2, \cdots),$$

where $[j,i]$, $[j]$, (j) are rational functions of the independent parameter m, namely [7]

$$(7) \quad [j,i] = -\frac{i}{j}\,\frac{4(j-1)i + 4j^2 + 4j - 2 - 4(i-j+1)m + m^2}{2(4j^2-1) - 4m + m^2},$$

$$(8) \quad [j] = -\frac{3}{16j^2}\,\frac{4j^2 - 8j - 2 - 4(j+2)m - 9m^2}{2(4j^2-1) - 4m + m^2},$$

$$(9) \quad (j) = -\frac{3}{16j^2}\,\frac{20j^2 - 16j + 2 - 4(5j-2)m + 9m^2}{2(4j^2-1) - 4m + m^2},$$

with the understanding that $j = \pm 1, \pm 2, \cdots$ but $i = 0, \pm 1, \pm 2, \cdots$.

Accordingly, the problem requires a proof to the effect that the infinite system (6), (5) of non-linear and non-recursive equations defines the infinitely many unknown functions (2), and that, for a fixed value (3) under consideration (and, in particular, for the fixed value (4) of m) the series (1) converge to functions $x = x(t)$, $y = y(t)$ which have second derivatives. For then (1) is a solution of the equations of motion, since (6), (5) represent all of the conditions obtained by the method of undetermined Fourier coefficients. Needless to say, $x = x(t)$, $y = y(t)$ must turn out to be analytic functions of t, the equations of motion being analytic. Incidentally, the functions (2) of m also prove to be analytic.

---

[5] G. W. Hill, *loc. cit.*[1], p. 145. This condition of Hill, when written in the above form (5), has as its right-hand member the number κ, instead of m^2. Actually, $\kappa = m^2$, if one chooses units in the usual way. In fact, Hill's n' denotes the constant angular velocity of the synodic coördinate system, while his μ is the total mass (*loc. cit.*[1], p. 9). Furthermore, $m = n'/(n - n')$ and $\kappa = \mu/(n - n')^2$ is his notation (*loc. cit.*[1], p. 9). Hence $\kappa = m^2$, if one chooses, as usual, $n' = 1$ and $\mu = 1$.

[6] *Ibid.*, p. 135. For reasons which will become apparent in (16) below, Hill's $[j]/m^2$ and $(j)/m^2$ have been denoted by $[j]$ and (j) respectively, while Hill's $[j,i]$ is the same as the above $[j,i]$.

[7] *Ibid.*, pp. 134-135. The factor m^2 of Hill's $[j]$ and (j) has been dropped in accordance with the preceding footnote.

Since the given rational functions (7), (8), (9) of the independent parameter m do not contain the unknown functions (2), the portion (6) of the complete system of conditions (5), (6) is homogeneous in the $a_i = a_i(m)$ and can, therefore, only determine the ratios

$$(10) \quad b_i = a_i : a_0 \equiv a_i(m) : a_0(m) = b_i(m) \qquad (i = \pm 1, \pm 2, \cdots)$$

of the unknown functions (2). If the functions (10) are known, the functions (2) follow from the remaining condition (5). In fact, the expression on the left of (5) is an infinite trilinear form in the $a_i = a_i(m)$; so that, if the functions (10) of m are known, a_0 follows as a function of m by the extraction of a cubic root. (Actually, the exponent of the radicand in m turns out to be $\frac{2}{3}$, instead of being merely $\frac{1}{3}$; and this $\frac{2}{3}$ expresses the infinitesimal validity of Kepler's third law).

Accordingly, the infinite system (5), (6) for the unknown functions (2) can be replaced by the infinite system (6) for the unknown functions (10).

A prepared form of the equations of condition. It is seen from (7) that

$$(11) \qquad [j, j] = -1 \quad \text{and} \quad [j, 0] = 0 \qquad (j = \pm 1, \pm 2, \cdots)$$

for every m. Hence, (6) can be written as

$$(12) \quad a_0 a_j = \sum_{i=-\infty}^{+\infty}{}' [j, i] a_i a_{i-j} + m^2 [j] \sum_{i=-\infty}^{+\infty}{}'' a_i a_{-i+j-1} + m^2(j) \sum_{i=-\infty}^{+\infty}{}''' a_i a_{-i-j-1} + 2m^2[j] a_0 a_{j-1} + 2m^2(j) a_0 a_{-j-1}, \qquad (j = \pm 1, \pm 2, \cdots),$$

where, corresponding to (11), the prime of the first sum means that

$$(13_1) \qquad i \neq j \text{ and } i \neq 0 \text{ in } \sum_{i=-\infty}^{+\infty}{}', \qquad (j = \pm 1, \pm 2, \cdots),$$

while two primes denote that

$$(13_2) \qquad i \neq j - 1 \text{ and } i \neq 0 \text{ in } \sum_{i=-\infty}^{+\infty}{}'', \qquad (j = \pm 1, \pm 2, \cdots),$$

finally three primes that

$$(13_3) \qquad i \neq -j - 1 \text{ and } i \neq 0 \text{ in } \sum_{i=-\infty}^{+\infty}{}''', \qquad (j = \pm 1, \pm 2, \cdots).$$

Since neither of the indices k, l of an arbitrary term $\cdots a_k a_l$ of the three sums (13_1)-(13_3) occurring in (12) vanishes, the homogeneous equations (12) for the unknown ratios (10) appear in the form

$$(14)\quad b_j = \sum_{i=-\infty}^{+\infty}{}' [j,i]b_ib_{i-j} + m^2[j]\sum_{i=-\infty}^{+\infty}{}'' b_ib_{-i+j-1} + m^2(j)\sum_{i=-\infty}^{+\infty}{}''' b_ib_{-i-j-1} + 2m^2[j]b_{j-1} + 2m^2(j)b_{-j-1}, \qquad (j = \pm 1, \pm 2, \cdots),$$

if one divides (12) by a_0^2.

Now consider, instead of the ratios (10), the unknowns c_j which are defined by

$$(15)\quad b_j(m) \equiv b_j = m^{|j|+1}c_j, \text{ so that } c_j = c_j(m) \qquad (j = \pm 1, \pm 2, \cdots).$$

Then the j-th of the equations (14) appears, upon division by $m^{|j|+1}$, in the form

$$(16)\quad c_j = m\{\sum_{i=-\infty}^{+\infty}{}' [j,i]m^{\alpha_{ij}}c_ic_{i-j} + [j]\sum_{i=-\infty}^{+\infty}{}'' m^{\beta_{ij}}c_ic_{-i+j-1} + (j)\sum_{i=-\infty}^{+\infty}{}''' m^{\gamma_{ij}}c_ic_{-i-j-1} + 2[j]m^{\delta_j}c_{j-1} + 2(j)m^{\epsilon_j}c_{-j-1}\}, \qquad (j = \pm 1, \pm 2, \cdots),$$

where the exponents are integers; in fact, corresponding to (13_1)-(13_3),

$$(17_1)\qquad \alpha_{ij} = |i-j| + |i| - |j| \text{ in } \sum_{i=-\infty}^{+\infty}{}';$$

$$(17_2)\qquad \beta_{ij} = |i+1-j| + |i| - |j| + 2 \text{ in } \sum_{i=-\infty}^{+\infty}{}'';$$

$$(17_3)\qquad \gamma_{ij} = |i+1+j| + |i| - |j| + 2 \text{ in } \sum_{i=-\infty}^{+\infty}{}''';$$

while obviously

$$(18_1)\quad \delta_j = |j-1| - |j| + 1; \qquad (18_2)\quad \epsilon_j = |j+1| - |j| + 1.$$

Since $j = \pm 1, \pm 2, \cdots$, comparison of (17_ν) with (13_ν), where $\nu = 1, 2, 3$, shows that

(19) the integers α_{ij}, β_{ij}, γ_{ij}, δ_j, ϵ_j are non-negative,

the inequalities $\delta_j \geqq 0$ and $\epsilon_j \geqq 0$ being obvious from (18_1)-(18_2).

Estimates for the coefficient functions. In what follows, use will be made of the notation

$$(20)\qquad f^* \equiv f^*(m) = \sum_{n=0}^{\infty} |c_n| m^n, \text{ if } f \equiv f(m) = \sum_{n=0}^{\infty} c_n m^n$$

is a regular analytic function of the independent variable m in the vicinity of the origin $m = 0$.

Since $j \neq 0$, any of the infinitely many rational functions (7), (8), (9) is regular at $m = 0$; so that (20) defines $[j,i]^*$, $[j]^*$, $(j)^*$ as power series in m with real non-negative coefficients. Now, all these power series are con-

vergent in the circle $|m| \leqq 1$ of the complex m-plane. Furthermore, there exists a bound C which is independent of j and i and is such that

$$|[j,i]^*| < C\,|i/j|^2 + C\,|i/j| \text{ for } |m| \leqq 1 \tag{21}$$

and

$$|[j]^*| < C/j^2 \text{ and } |(j)^*| < C/j^2 \text{ for } |m| \leqq 1. \tag{22}$$

In fact, if $f(m)$ is regular analytic in a circle $|m| < R$, and if $|f(m)| < M$ in this circle, then, on using the notations (20), one has $|c_n| < M/R^n$ (Cauchy), and so, if $R > 1$,

$$|f^*(m)| < MR/(R-1) \text{ for } |m| \leqq 1.$$

Consequently, (21) is true if there exists two constants M and R such that $R > 1$ and all of the functions (7) are regular and less than M in absolute value in the circle $|m| < R$. Now, on dividing the numerator and denominator of (7) by $2(4j^2-1)$, one sees that M and $R > 1$ exist if the rational functions

$$\left\{1 - \frac{4m - m^2}{2(4j^2-1)}\right\}^{-1}, \text{ where } j = \pm 1, \pm 2, \cdots, \tag{23}$$

are regular and uniformly bounded in a circle $|m| < R$, where $R > 1$. Finally, this condition is satisfied by $R = 8/7$, since if $|m| < 1 + 1/7$, then

$$|4m - m^2| < (1 + 1/7)(4 + 1 + 1/7) < 6 \leqq 2(4j^2-1) \text{ for } j = \pm 1, \pm 2, \cdots.$$

This proves (21); while (22) follows in the same way from (8), (9) as (21) did from (7).

A consequence of (21) and (22) is that if the variables m; $c_1, c_{-1}, c_2, c_{-2}, \cdots$ are restricted by the inequalities

$$|m| \leqq 1; \quad |c_j| \leqq 1/j^2 \qquad (j = \pm 1, \pm 2, \cdots) \tag{24}$$

but are otherwise arbitrary, then the estimates

$$\sum_{i=-\infty}^{+\infty}{}' |[j,i]^*\, m^{\alpha_{ij}} c_i c_{i-j}| < B/j^2, \qquad (j = \pm 1, \pm 2, \cdots) \tag{25}$$

and

$$\begin{aligned} &|[j]^*| \sum_{i=-\infty}^{+\infty}{}'' |m^{\beta_{ij}} c_i c_{-i+j-1}| < B/j^2, \\ &|(j)^*| \sum_{i=-\infty}^{+\infty}{}''' |m^{\gamma_{ij}} c_i c_{-i-j-1}| < B/j^2, \end{aligned} \qquad (j = \pm 1, \pm 2, \cdots) \tag{26}$$

hold for a sufficiently large constant B which, in addition, can be chosen such that

$$\begin{aligned} &2\,|[j]^*\, m^{\delta_j} c_{j-1}| < B/j^2, \\ &2\,|(j)^*\, m^{\epsilon_j} c_{-j-1}| < B/j^2, \end{aligned} \qquad (j = \pm 1, \pm 2, \cdots). \tag{27}$$

In fact, it is clear from (19) and (21) that, in the domain (24), the expression on the left of (25) is less than the product of the constant C and of

$$\sum_{i=-\infty}^{+\infty}{}' i^2 j^{-2} \cdot 1 \cdot i^{-2}(i-j)^{-2} + \sum_{i=-\infty}^{+\infty}{}' \mid ij^{-1} \mid \cdot 1 \cdot i^{-2}(i-j)^{-2}.$$

Since the first of these sums can be written as

$$\sum_{i=-\infty}^{+\infty}{}' j^{-2}(i-j)^{-2} = j^{-2} \cdot \text{const.},$$

and the second as

$$\sum_{i=-\infty}^{+\infty}{}' \mid ji \mid^{-1}(i-j)^{-2} < \mid j \mid^{-1} \sum_{i=-\infty}^{+\infty}{}' (i-j)^{-2} < \mid j \mid^{-1} \cdot \text{const.} \sum_{i=0}^{+\infty} (i + \mid j \mid)^{-2}$$

$$= \mid j \mid^{-1} \cdot \text{const.} \sum_{i=\mid j \mid}^{+\infty} i^{-2} < \mid j \mid^{-1} \cdot \text{Const.} \mid j \mid^{-1} = \text{Const.}/j^2,$$

the proof of (25) is complete.

Furthermore, (19) and (22) show that, in the domain (24), the expression on the left of the first of the inequalities (26) is less than

$$Cj^{-2} \sum_{i=-\infty}^{+\infty}{}'' i^{-2}(-i+j-1)^{-2} < 2Cj^{-2} \sum_{i=1}^{+\infty} i^{-2} = \text{const.}/j^2.$$

This proves the first of the inequalities (26), while the second follows in the same way.

Finally, (27) is clear from (19), (22) and (24).

The existence theorem. Without regard to questions of convergence, let

$$\phi_n \equiv \phi_n(z_0, z_1, \cdots) = \sum \cdots \sum a_{i_1 \ldots i_n} z_{i_1} \cdots z_{i_n},$$

where the n non-negative integers $i_1, \cdots, i_n$ need not be distinct, denote a form of degree n in the infinitely many variables $z_0, z_1, \cdots$, and suppose that the terms of this n-fold sum have been contracted completely, i. e. that the n equalities $i_1 = j_1, \cdots, i_n = j_n$ always imply the vanishing of at least one of the two constant coefficients $a_{i_1 \ldots i_n}$, $a_{j_1 \ldots j_n}$. On this assumption, associate with ϕ_n the unique form

$$\phi^*_n \equiv \phi^*_n(z_0, z_1, \cdots) = \sum \cdots \sum \mid a_{i_1 \ldots i_n} \mid z_{i_1} \cdots z_{i_n}$$

(so that $\phi_n \equiv \phi^*_n$ if and only if all $a \geqq 0$). And let f^* denote the unique power series

$$f^* \equiv f^*(z_0, z_1, \cdots) = \sum_{n=0}^{\infty} \phi^*_n, \text{ if } f \equiv f(z_0, z_1, \cdots) = \sum_{n=0}^{\infty} \phi_n$$

is a power series in the infinitely many variables $z_0, z_1, \cdots$; so that, corresponding to (20),

$$f^* = \sum_{n=0}^{\infty} \Sigma \cdots \Sigma \mid a_{i_1 \ldots i_n} \mid z_{i_1} \cdots z_{i_n}, \text{ if } f = \sum_{n=0}^{\infty} \Sigma \cdots \Sigma\, a_{i_1 \ldots i_n} z_{i_1} \cdots z_{i_n}.$$

Let there be given an infinite sequence

$$f_1(z; y_1, y_2, \cdots), \cdots, \qquad f_k(x; y_1, y_2, \cdots), \cdots \tag{28}$$

of such power series $f(z_0, z_1, z_2, \cdots)$, where $z_0 = x$; $z_1 = y_1$, $z_2 = y_2, \cdots$. Suppose that there exist two positive constants and two sequences of positive constants, say

$$\alpha; \mu \text{ and } B_1, \cdots, B_k, \cdots; M_1, \cdots, M_k, \cdots \tag{29}$$

such that, on the one hand,

$$B_k/M_k \geqq \mu > 0, \qquad (k = 1, 2, \cdots) \tag{30}$$

and, on the other hand, the sequence of power series (28) satisfies the conditions

$$f^*_k(\alpha; B_1, B_2, \cdots) \leqq M_k, \qquad (k = 1, 2, \cdots) \tag{31}$$

(which, in particular, imply that each of the power series (28) is absolutely convergent in the domain

$$\mid x \mid \leqq \alpha; \quad \mid y_1 \mid \leqq B_1, \cdots, \mid y_k \mid \leqq B_k, \cdots \tag{32}$$

of infinitely many dimensions).

Now (cf. *loc. cit.*[2] and *loc. cit.*[3]), if the power series (28) in infinitely many variables are such that the conditions (30), (31) are satisfied by positive constants (29), then the infinite implicit system of equations

$$y_k = x f_k(x; y_1, y_2, \cdots), \qquad (k = 1, 2, \cdots) \tag{33}$$

has in a sufficiently small fixed circle about the origin of the complex x-plane, namely at least in the circle

$$\mid x \mid < \min(\alpha, \mu) \tag{34}$$

which is independent of k, exactly one regular analytic solution $y_k = y_k(x)$. Furthermore,

$$\mid y_k(x) \mid < B_k \text{ for } \mid x \mid < \min(\alpha, \mu), \qquad (k = 1, 2, \cdots). \tag{35}$$

Finally, every $y_k(x)$ is real for real x if the coefficients of the power series (28) in infinitely many variables are real.

The proof of this theorem (*loc. cit.*[2] and *loc. cit.*[3]) is such as to supply not only the existence of the regular functions $y_k(x)$ and their estimates (35) but also convenient appraisals of the errors which are committed if one only carries out the process of the method of undetermined coefficients to a given stage.

Application of the existence theorem. Now, this existence theorem is applicable to the problem at hand, if one identifies

$$m;\ c_1, c_{-1}, c_2, c_{-2}, \cdots \text{ with } x;\ y_1, y_2, y_3, y_4, \cdots$$

respectively. In fact, the rational functions (7), (8), (9) of m are regular at $m = 0$ and can, therefore, be developed according to non-negative powers of $m = x$. Hence it is clear from (19) that the system (16) for the unknown functions $c_j = c_j(m)$ can be written in the form

$$c_j = mF_j(m; c_1, c_{-1}, c_2, c_{-2}, \cdots) \qquad (j = \pm 1, \pm 2, \cdots) \tag{36}$$

where the F_j are given power series in the infinitely many variables m; c_j and have real coefficients. Furthermore, there exists a constant $A > 0$ such that

$$F^*_j(1; 1^{-2}, 1^{-2}, 2^{-2}, 2^{-2}, \cdots) < Aj^{-2} \qquad (j = \pm 1, \pm 2, \cdots). \tag{37}$$

In fact, since (25), (26), (27) hold on the assumption (24), one can choose $A = 5B$, the function F_j in (36) being, by its definition, the function $\{\ \}$ in (16).

On identifying (36) with (33), one sees from (37) that the conditions of the existence theorem, namely (31) and (30), are satisfied by

$$\alpha = 1;\ B_j = 1/j^2,\ M_j = A/j^2 \qquad (j = \pm 1, \pm 2, \cdots)$$

since (30) is then satisfied by $\mu = 1/A$. Thus, the radius of the circle (34), where $x = m$, becomes $\min(1, 1/A)$; while (35) states that $|c_j(m)| < 1/j^2$, hence $|c_j(m)| < 1$, in this circle. Furthermore, $c_j(0) = 0$, as seen by placing $m = 0$ in (36).

Since (36) is identical with (16), while (16) is identical with (14) in virtue of (15), it follows that the system (14) has in the domain

$$|m| < \lambda, \text{ where } \lambda = \min(1, 1/A), \qquad (\beta < 1), \tag{38}$$

exactly one regular analytic solution (10), and that the Taylor series of the infinitely many functions $b_j(m)$ are of the form

$$b_j(m) = m^{|j|+2} \sum_{n=0}^{\infty} \beta_{jn} m^n \qquad (j = \pm 1, \pm 2, \cdots), \tag{39}$$

where the constants β_{jn} are real, finally that the power series (29) are in their common domain (38) of convergence such that

$$|b_j(m)| < |m|^{|j|}, \qquad (j = \pm 1, \pm 2, \cdots;\ |m| < \lambda < 1). \tag{40}$$

Now, (14) is, in virtue of (10), equivalent to (6), while (5) becomes a

cubic condition for the single function $a_0 = a_0(m)$, since the ratios (10) are the known functions (39). Finally, direct substitutions readily show that the unique real analytic function $a_0(m)$ which is thus defined by (5) does not vanish for small $m \gtrless 0$.

This completes the treatment of Hill's desideratum concerning (1). In fact, $a_0 = a_0(m)$ being a known function of m, the coefficients (2) of (1) follow from (10) and (39) uniquely, while (40) assures that the series (2) are Fourier series of (analytic) periodic functions of t.

The numerical situation. It should be mentioned that, on assuming in a formal way that $b_j(m)$ is a power series, it easily follows [8] from (6) or (14) without any reference to an existence theorem regarding (33), that $b_j(m)$ must be divisible by $m^{|j|}$. Needless to say, this does not dispose of the necessity of an existence and convergence proof, although $|m| < 1$. In fact, the usual formal treatment [9] of (6) does not take into account the possibility that not only (40) might not hold on a circle (38) which is independent of j but also that some or all of the power series (39) might be divergent for every $m \neq 0$, although $b_j(m)$ is, in a formal way, divisible by $m^{|j|}$. Actually, it is clear a priori that if the periodic solution (1) exists at all for a fixed m which does not correspond to a collision path, then $|a_j| < \text{const.}\ \vartheta^{|j|}$ for a suitable positive $\vartheta < 1$, since the periodic functions (1) must be analytic in the real variable t; so that (40) is by no means surprising, if the existence of (1) or (2) is assumed.

Loc. cit.[2], an existence theorem concerning (33) was applied to (14) instead of to (16), so that the numerical advantages of Hill's equations of condition were not completely utilized. Furthermore, the existence theorem applied *loc. cit.*[2] was a particular case of the one used above, since, instead of the general condition (30) used above, it was assumed that all B_k and all M_k are equal. This necessitated (*loc. cit.*[2]) the use of an absolute constant connected with a fact which is fundamental in Riemann's theory of trigonometrical series, namely with the fact that

$$\sum_{i=-\infty}^{+\infty}{}' \frac{1}{i^2(i-j)^2} = O\left(\frac{1}{j^2}\right) \text{ as } j \to \pm\infty.$$

All of this has made the numerical situation quite unfavorable, although it has allowed the proof of the fact [10] that the process is valid in a circle (38) whose radius is not less than $\frac{1}{12}$. While this circle contains the value (4)

---

[8] Cf. H. Poincaré, *Leçons de mécanique céleste*, vol. 2, part 2 (1909), pp. 35-36.

[9] Cf. H. Poincaré, *loc. cit.*[8], pp. 34-39.

[10] A. Wintner, "Über die Konvergenzfragen der Mondtheorie," *Mathematische Zeitschrift*, vol. 30 (1929), pp. 211-227.

belonging to the case of the Moon, Hill's numerical evidence indicates that the convergence must be much stronger than what one would expect of the closeness of the value (4) to the bound $\frac{1}{12} = 0.08333\cdots$, if this bound were sharp. Now, for reasons explained before, the method of the present paper leads to bounds much better than $\frac{1}{12}$. The detailed numerical work to this effect is in progress.

Of course, a bound of this type, no matter how sharp, is necessarily unable to answer the important question as to the "natural termination" (Strömgren) of Hill's family of periodic orbits.[11] All that follows from the general theory is that such a "dynamical termination" exists.[12] As far as the problem of convergence in the vicinity of $m = \pm 0$ is concerned, it is clear that complex singularities in m, which have no significance for the problem of termination, can act unfavorably.

Correspondingly, it is, according to Hill (*loc. cit.*[1]), convenient to carry out the expansions, not according to the powers of m but rather according to the powers of a linear function $\tilde{m} = \tilde{m}(m)$ of m which vanishes with m and is chosen so as to transform those complex singularities in m which are formally indicated by (23) into positions which are not near the real axis. Clearly, this amounts to what is called, in the theory of divergent series, an Euler transformation. For instance, if m is positive and small, one improves on the closeness of the complex singularities of the rational functions (23) for $j = \pm 1, \pm 2, \cdots$, if one replaces m by the expansion parameter

$$m' = m/(1 + m) \tag{41}$$

of Delaunay's lunar theory. It is easy to rewrite the existence proof of the present paper from m to this m' or to Hill's $\tilde{m} = \tilde{m}(m)$, and then to calculate the corresponding bounds, thus improved. This numerical work also is in progress.

Needless to say, a transformation of the type (41) is well adapted only to the case of direct paths, i. e., to the region $(m > 0)$ of the Moon. In view of a result of Birkhoff,[13] the initial behavior of the retrograde paths $(m < 0)$, which are not excluded in (38), is a much easier problem than the corresponding behavior of the direct paths.

THE JOHNS HOPKINS UNIVERSITY.

---

[11] Cf. A. Wintner, "Almost periodic functions and Hill's theory of lunar perigee," *American Journal of Mathematics*, vol. 59 (1937), pp. 795-802, footnote 12 on p. 798.

[12] A. Wintner, "Beweis des E. Strömgrenschen dynamischen Abschlussprincips," *Mathematische Zeitschrift*, vol. 34 (1931), pp. 321-349. Cf. G. D. Birkhoff, *Sur le problème restreint des trois corps*, Second mémoire, chap. III, § 5, *Annali della R. Scuola Normale Superiore di Pisa*, serie 2, vol. 5 (1936).

[13] G. D. Birkhoff, *loc. cit.*[4], § 18. Cf. also A. Wintner, *loc. cit.*[4], § 8–§ 9.

THE JOHNS HOPKINS PRESS · BALTIMORE

American Journal of Mathematics. Edited by E. T. Bell, Abraham Cohen, T. H. Hildebrandt, F. D. Murnaghan and J. F. Ritt. Quarterly. 8vo. Volume LX in progress. $7.50 per volume. (Foreign postage, fifty cents.)

American Journal of Philology. Edited by Tenney Frank with the coöperation of H. Cherniss, K. Malone, B. D. Meritt, and D. M. Robinson. Quarterly. 8vo. Volume LIX in progress. $5 per volume. (Foreign postage, twenty-five cents.)

Biologia Generalis. (International Journal of Biology). Founded by Leopold Löhner, Graz; Raymond Pearl, Baltimore, and Vladislav Růžička, Prague. It is now edited by O. Abel, L. Adametz, O. Porsch, C. Schwarz, J. Versluys and R. Wasicky of Vienna. 8vo. Volume XIV in progress.

Bulletin of the Institute of the History of Medicine. Edited by Henry E. Sigerist. Monthly except August and September. Volume VI in progress. 8vo. Subscription $5 per year. (Foreign postage, fifty cents.)

Bulletin of the Johns Hopkins Hospital. Edited by James Bordley, III. Monthly. Volume LXII in progress. 8vo. Subscription $6 per year. (Foreign postage, fifty cents.)

Comparative Psychology Monographs. Roy M. Dorcus, Managing Editor. 8vo. Volume XV in progress. $5 per volume.

Hesperia. Edited by William Kurrelmeyer and Kemp Malone. 8vo. Twenty-nine numbers have appeared.

Human Biology: a record of research. Raymond Pearl, Editor. Quarterly. 8vo. Volume X in progress. $5 per volume. Foreign postage, thirty-five cents.)

Johns Hopkins Studies in Romance Literatures and Languages. H. C. Lancaster, Editor. 8vo. Forty-two numbers have been published.

Johns Hopkins University Circular, including the President's Report and Catalogue of the School of Medicine. Ten times yearly. 8vo. $1 per year.

Johns Hopkins University Studies in Archaeology. David M. Robinson, Editor. 8vo. Twenty-five volumes have appeared.

Johns Hopkins University Studies in Education. Florence E. Bamberger, Editor. 8vo. Twenty-six numbers have appeared.

Johns Hopkins University Studies in Geology. Edward B. Mathews, Editor. 8vo. Twelve numbers have been published.

Johns Hopkins University Studies in Historical and Political Science. Under the direction of the Departments of History, Political Economy and Political Science. 8vo. Volume LVI in progress. $5 per volume.

Modern Language Notes. Edited by H. C. Lancaster, W. Kurrelmeyer, R. D. Havens, K. Malone, H. Spencer and C. S. Singleton. Eight times yearly. 8vo. Volume LIII in progress. $5 per volume. (Foreign postage, fifty cents.)

Reprint of Economic Tracts. J. H. Hollander, Editor. Fifth series in progress. Price $4.

Terrestrial Magnetism and Atmospheric Electricity. Founded by Louis A. Bauer; conducted by J. A. Fleming with the coöperation of eminent investigators. Quarterly. 8vo. Volume XLIII in progress. $3.50 per volume.

Walter Hines Page School of International Relations. Six volumes have been published.

A complete list of publications will be sent upon request

THE THEORY OF GROUP REPRESENTATIONS

By Francis D. Murnaghan

We have attempted to give a quite elementary and self-contained account of the theory of group representations with special reference to those groups (particularly the symmetric group and the rotation group) which have turned out to be of fundamental significance for quantum mechanics (especially nuclear physics). We have devoted particular attention to the theory of group integration (as developed by Schur and Weyl); to the theory of two-valued or spin representations; to the representations of the symmetric group and the analysis of their direct products; to the crystallographic groups; and to the Lorentz group and the concept of semi-vectors (as developed by Einstein and Mayer). —*Extract from Preface.*

Ready in November. Price $5.00

NUMERICAL MATHEMATICAL ANALYSIS

By JAMES B. SCARBOROUGH

"A valuable feature of the book is the excellent collection of examples at the end of each chapter. . . . The book has many admirable features. The explanations and derivations of formulae are given in detail. . . . The author has avoided introducing new and complicated notations which, although they may conduce to brevity, are a serious stumbling block to the reader. The typography and paper are excellent."

—*American Mathematics Monthly.*

430 pages, 25 figures, crown 8vo, buckram, $5.50

TABLES OF $\sqrt{1-r^2}$ AND $1-r^2$ FOR USE IN PARTIAL CORRELATION AND IN TRIGONOMETRY

By JOHN RICE MINER, Sc. D.

These tables fill a want long felt by practical workers in all branches of statistics. Everyone who uses the method of correlation has wished for tables from which the probable error of a coefficient of correlation could be obtained *with accuracy*. Similar tables to this have existed on a small scale, but never before have there been available tables of $1-r^2$ and $\sqrt{1-r^2}$ to 6 places of decimals, and 4 places in the argument. Not only are these tables of great usefulness in getting the probable error of a correlation coefficient, but also they have what will perhaps be their chief value in the calculations involved in the method of partial or net correlation. It is safe to say that these tables reduce the labor involved in this widely used statistical method by at least one-half.

50 pages, 8vo, cloth, $1.50

THE JOHNS HOPKINS PRESS · BALTIMORE